



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/56, 9/24, C12Q 1/68, C07K 16/40, A61K 38/47	A1	(11) International Publication Number: WO 97/19177 (43) International Publication Date: 29 May 1997 (29.05.97)
(21) International Application Number: PCT/AU96/00747 (22) International Filing Date: 22 November 1996 (22.11.96) (30) Priority Data: PN 6748 23 November 1995 (23.11.95) AU (71) Applicant (for all designated States except US): WOMEN'S AND CHILDREN'S HOSPITAL [AU/AU]; 72 King William Road, North Adelaide, S.A. 5006 (AU). (72) Inventors; and (75) Inventors/Applicants (for US only): HOPWOOD, John, Joseph [AU/AU]; 2 Monarto Court, Stonyfell, S.A. 5066 (AU). SCOTT, Hamish, Steele [AU/CH]; 43, rue des Bains, CH-1205 Genève (CH). WEBER, Birgit [DE/AU]; 5/15 Botanic Street, Hackney, S.A. 5069 (AU). BLANCH, Lianne [AU/AU]; 469 The Esplanade, Grange, S.A. 5022 (AU). ANSON, Donald, Stewart [GB/AU]; 12 Ross Street, Thebarton, S.A. 5031 (AU). (74) Agents: SLATTERY, John, Michael et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).		(81) Designated States: AU, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: SYNTHETIC MAMMALIAN α -N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME (57) Abstract The present invention relates generally to mammalian α -N-acetylglucosaminidase and to genetic sequences encoding same and to their use in the investigation, diagnosis and treatment of subjects suspected of or suffering from α -N-acetylglucosaminidase deficiency.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

SYNTHETIC MAMMALIAN α -N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME

5

FIELD OF THE INVENTION

The present invention relates generally to mammalian α -N-acetylglucosaminidase and to genetic sequences encoding same and to the use of these in the investigation, diagnosis and treatment of subjects suspected of or suffering from α -N-acetylglucosaminidase deficiency.

10

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description. Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined following the bibliography.

15

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

20

BACKGROUND TO THE INVENTION

The increasing sophistication of recombinant DNA technology is greatly facilitating the efficacy of many commercially important industries including areas of medical and pharmaceutical research and development. The ability to purify native proteins and subsequently clone genetic sequences encoding these proteins is an important first step in the development of a range of therapeutic and diagnostic procedures. However, practitioners have faced many difficulties in purifying target molecules to an extent sufficient to determine amino acid sequences to permit the development of oligonucleotide probes to assist in the cloning of genetic sequences encoding the target molecules.

30

Such difficulties have been particularly faced in the research and development of lysosomal enzymes. An important lysosomal enzyme is α -N-acetylglucosaminidase (EC 2.1.50). This enzyme acts as a exoglycosidase in lysosomes to hydrolyse the terminal α -N-acetylglucosamine residues present at the non-reducing terminus of fragments of heparan sulphate and heparin (Hopwood, 1989). A deficiency in this lysosomal hydrolase is responsible for the pathogenesis of Sanfilippo B (Mucopolysaccharidosis type IIIB [MPS-IIIB]) syndrome (von-Figura and Kresse, 1972; O'Brien, 1972). This is an autosomal recessive disorder of glycosaminoglycan catabolism leading to storage and excretion of excessive amounts of heparan sulphate and a variety of clinical phenotypes, but classically presenting with progressive mental retardation in conjunction with skeletal deformities (McKusick and Neufeld, 1983).

There is a need, therefore, to purify α -N-acetylglucosaminidase and to clone genetic sequences encoding same to permit development of a range of therapeutic and diagnostic procedures to assist in the diagnosis and treatment of disease conditions arising from α -N-acetylglucosaminidase deficiency.

20 SUMMARY OF THE INVENTION

One aspect of the invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes or is complementary to a sequence which encodes a mammalian α -N-acetylglucosaminidase or fragment or derivative thereof.

25 A second aspect of the invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which is capable of hybridising under at least low stringency conditions to a nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or a complementary strand or a homologue, analogue or derivative thereof.

30 Another aspect of the invention is directed an isolated nucleic acid molecule which is at least 40% identical to the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID

- 3 -

NO:3 or to a complementary strand thereof or a homologue, analogue or derivative thereof.

5 A further aspect of the present invention provides a nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a polypeptide capable of hydrolysing the terminal α -N-acetylglucosamine residues present at the non-reducing terminus of fragments of heparan sulphate and heparin residues and wherein said nucleotide sequence is capable of hybridising under low stringency conditions to the nucleotide sequence set forth in SEQ ID NO:1.

10

A further aspect of the invention is directed to a genetic construct comprising a sense molecule, for the expression or over-expression of α -N-acetylglucosaminidase in prokaryotic or eukaryotic cells.

15 A further aspect of the present invention is directed to synthetic α -N-acetylglucosaminidase or like molecule.

A further aspect of the invention contemplates antibodies to α -N-acetylglucosaminidase and preferably synthetic α -N-acetylglucosaminidase or a like molecule.

20

In still yet another aspect of the present invention there is contemplated a method of diagnosing a mutation or other aberrations in the α -N-acetylglucosaminidase gene in a human or animal patient.

25 Another aspect contemplates a method of treating patients suffering from α -N-acetylglucosaminidase deficiency, such as in MPS-IIIB, said method comprising administering to said patient an effective amount of α -N-acetylglucosaminidase or active like form thereof.

30 Another aspect of the present invention is directed to a pharmaceutical composition comprising a recombinant mammalian α -N-acetylglucosaminidase or an active fragment

or derivative thereof and one or more pharmaceutically acceptable carriers and/or diluents.

5 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a photographic representation of α -N-acetylglucosaminidase purified from human placenta following SDS/PAGE. Lane 1: M_r standards (kDa); Lanes 2 and 3 : purified α -N-acetylglucosaminidase from human placenta. Lane 4 and 5, bovine serum albumin.

10

Figure 2 is a photographic representation of an SDS/polyacrylamide gel showing the molecular weights of recombinant α -N-acetylglucosaminidase polypeptides produced in CHO cells before (-) and after (+) PNGase F digestion. The 50 mM NaCl and 75 mM NaCl fractions are indicated. Molecular weights of α -N-acetylglucosaminidase polypeptides are indicated on the left of the figure. Molecular weights of marker proteins are indicated on the right hand side of the figure (lane 5).

15

Single and three letter abbreviations of conventional amino acid residues as used herein are defined in Table 1.

20

Suitable amino acid substitutions referred to herein are defined in Table 2.

Codes for non-conventional amino acid residues as used herein are defined in Table 3.

- 5 -

TABLE 1

	Amino Acid	Three-letter Abbreviation	One-letter Symbol
5	Alanine	Ala	A
	Arginine	Arg	R
	Asparagine	Asn	N
	Aspartic acid	Asp	D
10	Cysteine	Cys	C
	Glutamine	Gln	Q
	Glutamic acid	Glu	E
	Glycine	Gly	G
	Histidine	His	H
15	Isoleucine	Ile	I
	Leucine	Leu	L
	Lysine	Lys	K
	Methionine	Met	M
	Phenylalanine	Phe	F
20	Proline	Pro	P
	Serine	Ser	S
	Threonine	Thr	T
	Tryptophan	Trp	W
	Tyrosine	Tyr	Y
25	Valine	Val	V
	Any residue	Xaa	X

- 6 -

TABLE 2
Suitable residues for amino acid substitutions

	<u>Original Residue</u>	<u>Exemplary Substitutions</u>
5	Ala	Ser
	Arg	Lys
	Asn	Gln; His
	Asp	Glu
	Cys	Ser
10	Gln	Asn
	Glu	Asp
	Gly	Pro
	His	Asn; Gln
	Ile	Leu; Val
15	Leu	Ile; Val
	Lys	Arg; Gln; Glu
	Met	Leu; Ile
	Phe	Met; Leu; Tyr
	Ser	Thr
20	Thr	Ser
	Trp	Tyr
	Tyr	Trp; Phe
	Val	Ile; Leu

25

- 7 -

TABLE 3

	Non-conventional amino acid	Code	Non-conventional amino acid	Code
5	α -aminobutyric acid	Abu	L-N-methylalanine	Nmala
	α -amino- α -methylbutyrate	Mgab	L-N-methylarginine	Nmarg
	aminocyclopropane- carboxylate	Cpro	L-N-methylasparagine	Nmasn
			L-N-methylaspartic acid	Nmasp
10	aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
	aminonorbornyl- carboxylate	Norb	L-N-methylglutamine	Nmgln
			L-N-methylglutamic acid	Nmglu
	cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
	cyclopentylalanine	Cpen	L-N-methylisoleucine	Nmile
15	D-alanine	Dal	L-N-methylleucine	Nmleu
	D-arginine	Darg	L-N-methyllysine	Nmlys
	D-aspartic acid	Das	L-N-methylmethionine	Nmmet
	D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
	D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20	D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
	D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
	D-isoleucine	Dile	L-N-methylproline	Nmpro
	D-leucine	Dleu	L-N-methylserine	Nmser
	D-lysine	Dlys	L-N-methylthreonine	Nmthr
25	D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
	D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
	D-phenylalanine	Dphe	L-N-methylvaline	Nmval
	D-proline	Dpro	L-N-methylethylglycine	Nmetg
	D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30	D-threonine	Dthr	L-norleucine	Nle
	D-tryptophan	Dtrp	L-norvaline	Nva

- 8 -

	D-tyrosine	Dtyr	α -methyl-aminoisobutyrate	Maib
	D-valine	Dval	α -methyl- γ -aminobutyrate	Mgab
	D- α -methylalanine	Dmala	α -methylcyclohexylalanine	Mchexa
	D- α -methylarginine	Dmarg	α -methylcyclopentylalanine	Mcpen
5	D- α -methylasparagine	Dmasn	α -methyl- α -naphthylalanine	Manap
	D- α -methylaspartate	Dmasp	α -methylpenicillamine	Mpen
	D- α -methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D- α -methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D- α -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D- α -methylisoleucine	Dmile	N-amino- α -methylbutyrate	Nmaabu
	D- α -methylleucine	Dmleu	α -naphthylalanine	Anap
	D- α -methyllysine	Dmlys	N-benzylglycine	Nphe
	D- α -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D- α -methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D- α -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D- α -methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D- α -methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D- α -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D- α -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D- α -methyltyrosine	Dmtty	N-cyclodecylglycine	Ncdec
	D- α -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl) glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl) glycine	Nbhe

- 9 -

	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl) glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
5	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl) glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl) glycine	Nhtrp
	D-N-methyllysine	Dnmlys	N-methyl- γ -aminobutyrate	Nmgabu
10	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmt
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpn
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
15	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyl- α -naphthylalanine	Nmanap
	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ -aminobutyric acid	Gabu	N-(<i>p</i> -hydroxyphenyl)glycine	Nhtyr
20	L- <i>t</i> -butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L- α -methylalanine	Mala
	L- α -methylarginine	Marg	L- α -methylasparagine	Masn
	L- α -methylaspartate	Masp	L- α -methyl- <i>t</i> -butylglycine	Mtbug
25	L- α -methylcysteine	Mcys	L-methylethylglycine	Metg
	L- α -methylglutamine	Mgln	L- α -methylglutamate	Mglu
	L- α -methylhistidine	Mhis	L- α -methylhomo phenylalanine	Mhphe
	L- α -methylisoleucine	Mile	N-(2-methylthioethyl) glycine	Nmet
30	L- α -methylleucine	Mleu	L- α -methyllysine	Mlys

- 10 -

	L- α -methylmethionine	Mmet	L- α -methylnorleucine	Mnle
	L- α -methylnorvaline	Mnva	L- α -methylornithine	Morn
	L- α -methylphenylalanine	Mphe	L- α -methylproline	Mpro
	L- α -methylserine	Mser	L- α -methylthreonine	Mthr
5	L- α -methyltryptophan	Mtrp	L- α -methyltyrosine	Mtyr
	L- α -methylvaline	Mval	L-N-methylhomo	
			phenylalanine	Nmhpe
	N-(N-(2,2-diphenylethyl)		N-(N-(3,3-diphenylpropyl)	
	carbamylmethyl)glycine	Nnbhm	carbamylmethyl)glycine	Nnbhe
10	1-carboxy-1-(2,2-diphenyl-			
	ethylamino)cyclopropane	Nmbc		

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes, or are complementary to a sequence which encodes, a
5 mammalian α -N-acetylglucosaminidase or fragment or derivative thereof or its like molecule.

Preferably, the mammal is a human, livestock animal, companion animal, wild animal or laboratory test animal (e.g. rabbit, rat, mouse or guinea pig). Most preferably, the mammal
10 is a human. Conveniently, the α -N-acetylglucosaminidase is isolatable from the liver, kidney or placenta. However, the present invention extends to all mammalian α -N-acetylglucosaminidase enzymes and from any anatomical or cellular source and/or any biological fluid source, such as but not limited to plasma, serum, cell extract or lymph fluid.

15 Although a preferred embodiment of the present invention contemplates the use of human α -N-acetylglucosaminidase or genomic or recombinant (e.g. cDNA) genetic sequences encoding same in the investigation, diagnosis and/or treatment of human subjects (i.e. homologous system), one skilled in the art will appreciate that the enzyme or genetic
20 sequences encoding same from a non-human animal may also be useful. Such a heterologous system is encompassed by the present invention.

The term "nucleic acid molecule" as used herein shall be taken to refer to any RNA or DNA (eg. cDNA) molecule, whether single-stranded or double-stranded or in a linear or
25 covalently-closed form. The nucleic acid molecule may also be DNA corresponding to the entire genomic gene or a substantial portion thereof or a fragment or derivative thereof.

The nucleic acid molecule of the present invention may constitute solely the nucleotide sequence encoding α -N-acetylglucosaminidase or a α -N-acetylglucosaminidase-like
30 molecule or may be part of a larger nucleic acid molecule. Accordingly, the present invention extends to the isolated genomic α -N-acetylglucosaminidase gene. The non-

- 12 -

translated sequences in a larger nucleic acid molecule may include vector, transcriptional and/or translational regulatory sequences, promoter, terminator, enhancer, replication or signal sequences or non-coding regions (eg intron sequences) of an isolated genomic gene.

5 Reference herein to a "gene" is to be taken in its broadest context and includes:

- (i) a classical genomic gene consisting of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (i.e. introns, 5'- and 3'- untranslated sequences);
- (ii) mRNA or cDNA corresponding to the coding regions (i.e. exons)
10 optionally comprising 5'- or 3'-untranslated sequences of the gene; or
- (iii) synthetic, amplified DNA fragments or other recombinant nucleic acid molecules produced *in vitro* and comprising all or a part of the coding region and/or 5'- or 3'-untranslated sequences of the gene.

15 The term "gene" is also used to describe synthetic or fusion molecules encoding all or part of a functional product. A functional product is one which comprises a sequence of nucleotides or is complementary to a sequence of nucleotides which encodes a functional polypeptide, in particular a polypeptide having the catalytic activity of α -N-acetylglucosaminidase or a homologue, analogue or derivative thereof.

20

For the present purpose, "homologues" of a nucleotide sequence shall be taken to refer to an isolated nucleic acid molecule which is substantially the same as the nucleic acid molecule of the present invention or its complementary nucleotide sequence, notwithstanding the occurrence within said sequence, of one or more nucleotide
25 substitutions, insertions, deletions, or rearrangements.

"Analogues" of a nucleotide sequence set forth herein shall be taken to refer to an isolated nucleic acid molecule which is substantially the same as a nucleic acid molecule of the present invention or its complementary nucleotide sequence, notwithstanding the
30 occurrence of any non-nucleotide constituents not normally present in said isolated nucleic acid molecule, for example carbohydrates, radiochemicals including

- 13 -

radionucleotides, reporter molecules such as, but not limited to DIG, alkaline phosphatase or horseradish peroxidase, amongst others.

"Derivatives" of a nucleotide sequence set forth herein shall be taken to refer to any
5 isolated nucleic acid molecule which contains significant sequence similarity to said
sequence or a part thereof. Generally, the nucleotide sequence of the present invention
may be subjected to mutagenesis to produce single or multiple nucleotide substitutions,
deletions and/or insertions. Nucleotide insertional derivatives of the nucleotide sequence
of the present invention include 5' and 3' terminal fusions as well as intra-sequence
10 insertions of single or multiple nucleotides or nucleotide analogues. Insertional
nucleotide sequence variants are those in which one or more nucleotides or nucleotide
analogues are introduced into a predetermined site in the nucleotide sequence of said
sequence, although random insertion is also possible with suitable screening of the
resulting product being performed. Deletional variants are characterised by the removal
15 of one or more nucleotides from the nucleotide sequence. Substitutional nucleotide
variants are those in which at least one nucleotide in the sequence has been removed and
a different nucleotide or nucleotide analogue inserted in its place.

Preferably, a homologue, analogue or derivative of an α -N-acetylglucosaminidase gene
20 according to any embodiments described herein, comprises a sequence of nucleotides of
at least 10 contiguous nucleotides derived from SEQ ID NO:1 or SEQ ID NO:3 or a
complementary strand thereof, wherein the sequence of said homologue, analogue or
derivative is at least 40% identical to SEQ ID NO:1 or SEQ ID NO:3 or a
complementary strand thereof or wherein said homologue, analogue or derivative is
25 capable of hybridising to said sequence under at least low stringency hybridisation
conditions.

For the purposes of nomenclature, the nucleotide sequence set forth in SEQ ID NO: 1
relates to the cDNA encoding the human α -N-acetylglucosaminidase enzyme.

30

The nucleotide sequence set forth in SEQ ID NO:3 relates to the genomic gene equivalent

- 14 -

of the cDNA encoding the human liver α -N-acetylglucosaminidase enzyme. Those skilled in the art will be aware that the specific exon sequences described in SEQ ID NO:3 correspond to the coding regions of the α -N-acetylglucosaminidase gene, said exon regions further comprising the entire open reading frame of the cDNA sequence set forth in SEQ ID NO:1, when aligned in a head-to-tail configuration. The intron sequences of SEQ ID NO:3, which correspond to non-coding regions of the gene which are spliced from the primary transcription product thereof, although not explicitly defined, may be readily deduced by those skilled in the art, when provided with the exon sequence data provided in the nucleotide sequence listing.

10

The nucleotide sequence of the present invention may correspond to the sequence of the naturally-occurring α -N-acetylglucosaminidase gene or may comprise a homologue, analogue or derivative thereof which contains single or multiple nucleotide substitutions, deletions and/or additions. All such homologues, analogue or derivatives encode α -N-acetylglucosaminidase or α -N-acetylglucosaminidase-like molecules or a homologue, analogue or derivative thereof as contemplated by the present invention. The length of the nucleotide sequence may vary from a few bases, such as in nucleic acid probes or primers, to a full length sequence.

20 The present invention is particularly directed to the nucleic acid in cDNA form and particularly when inserted into an expression vector. The expression vector may be replicable in a eukaryotic or prokaryotic cell and may either produce mRNA or the mRNA may be subsequently translated into α -N-acetylglucosaminidase or like molecule. Particularly preferred eukaryotic cells include CHO cells but may be in any other suitable mammalian cells or cell lines or non-mammalian cells such as yeast or insect cells.

In an alternative embodiment, the present invention provides a nucleic acid molecule comprising a sequence of nucleotides which encodes or are complementary to a sequence which encodes a polypeptide capable of hydrolysing the α -N-acetylglucosamine residues from the non-reducing terminus of heparan sulphate and heparin fragments and wherein said nucleotide sequence is capable of hybridising under at least low stringency conditions

30

- 15 -

to a nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or a homologue, analogue or derivative thereof.

5 A second aspect of the invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which is capable of hybridising under at least low stringency conditions to a nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or a complementary strand or a homologue, analogue or derivative thereof.

10 Preferably, hybridisation is possible under at least medium stringent conditions. More preferably, hybridisation is possible under high stringent conditions.

For the purposes of defining the level of stringency, reference can conveniently be made to Sambrook *et al* (1989) or Ausubel *et al* (1987) which are herein incorporated by reference.

15

A low stringency is defined herein as being a hybridisation and/or wash carried out in 4-6X SSC/0.1-0.5% w/v SDS at 37-45°C for 2-3 hours. A medium stringency hybridisation and/or wash is carried out in 1-4X SSC/0.25-0.5% w/v SDS at $\geq 45^{\circ}\text{C}$ for 2-3 hours and a high stringency hybridisation and/or wash is carried out 0.1-1X SSC/0.1% w/v SDS at
20 60°C for 1-3 hours.

Alternative conditions of stringency may be employed to those specifically recited herein. Generally, the stringency is increased by reducing the concentration of SSC buffer, and/or increasing the concentration of SDS and/or increasing the temperature of the
25 hybridisation and/or wash. Those skilled in the art will be aware that the conditions for hybridisation and/or wash may vary depending upon the nature of the hybridisation membrane or the type of hybridisation probe used. Conditions for hybridisations and washes are well understood by one normally skilled in the art. For the purposes of clarification of parameters affecting hybridisation between nucleic acid molecules,
30 reference is found in pages 2.10.8 to 2.10.16. of Ausubel *et al.* (1987), which is herein incorporated by reference.

- 16 -

Those skilled in the art will be aware that the nucleotide sequences set forth in SEQ ID NO:1 and SEQ ID NO:3 may be used to isolate the corresponding genes from other human tissues or alternatively, from the tissues or cells of other species, without undue experimentation. Means for the isolated of such related sequences will also be known
5 to those skilled in the art, for example nucleic acid hybridisation, polymerase chain reaction, antibody screening of expression libraries, functional screening of expression libraries, or complementation of mutants, amongst others. The present invention is not to be limited by the source from which the specific gene sequences described herein have been isolated or by the means used to isolate said sequences.

10

In one embodiment, a related genetic sequence comprising genomic DNA, or mRNA, or cDNA is contacted with a hybridisation effective amount of a genetic sequence which encodes α -N-acetylglucosaminidase, or its complementary nucleotide sequence or a homologue, analogue, derivative or functional part thereof, and then said hybridisation
15 is detected using a suitable detection means.

The related genetic sequence may be in a recombinant form, in a virus particle, bacteriophage particle, yeast cell, animal cell, or a plant cell. Preferably, the related genetic sequence originates from an animal species or a human. More preferably, the
20 related genetic sequence originates from a human.

Preferably, the genetic sequence which encodes α -N-acetylglucosaminidase (i.e probe or latter genetic sequence) comprises a sequence of nucleotides of at least 10 nucleotides, more preferably at least 20 nucleotides, even more preferably at least 50 nucleotides and
25 even still more preferably at least 100 nucleotides derived from the nucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO:3 or a complementary sequence or a homologue, analogue or derivative thereof.

Preferably, the detection means is a reporter molecule capable of giving an identifiable
30 signal (e.g. a radioisotope such as ^{32}P or ^{35}S or a biotinylated molecule) covalently attached to the α -N-acetylglucosaminidase probe.

- 17 -

In an alternative embodiment, the detection means is a polymerase chain reaction. According to this embodiment, two opposing non-complementary nucleic acid "primer molecules" of at least 10 nucleotides in length, more preferably at least 20 nucleotides in length, derived from the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 may be contacted with a nucleic acid "template molecule" and specific nucleic acid molecule copies of the template molecule amplified in a polymerase chain reaction.

The opposing primer molecules are selected such that they are each capable of hybridising to complementary strands of the same template molecule, wherein DNA polymerase-dependant DNA synthesis occurring from a first opposing primer molecule will be in a direction toward the second opposing primer molecule.

Accordingly, both primers hybridise to said template molecule such that, in the presence of a DNA polymerase enzyme, a cofactor and appropriate substrate, DNA synthesis occurs in the 5' to 3' direction from each primer molecule towards the position on the DNA where the other primer molecule is hybridised, thereby amplifying the intervening DNA.

Those skilled in the art are aware of the technical requirements of the polymerase chain reaction and are capable of any modifications which may be made to the reaction conditions. For example, of the polymerase chain reaction may be used in any suitable format, such as amplified fragment length polymorphism (AFLP), single-strand chain polymorphism (SSCP), amplification and mismatch detection (AMD), interspersed repetitive sequence polymerase chain reaction (IRS-PCR), inverse polymerase chain reaction (iPCR) and reverse transcription polymerase chain reaction (RT-PCR), amongst others, to isolate a related α -N-acetylglucosaminidase gene sequence or identify a mutation in an α -N-acetylglucosaminidase genetic sequence. Such variations of the polymerase chain reaction are discussed in detail by McPherson *et al* (1991), which is incorporated herein by reference. The present invention encompasses all such variations, the only requirement being that the final product of the reaction is an isolated nucleic acid molecule which is capable of encoding α -N-acetylglucosaminidase or a homologue,

- 18 -

analogue or derivative thereof.

In a preferred embodiment, the first primer molecule is preferably derived from the sense strand of a gene which encodes α -N-acetylglucosaminidase, in particular from the
5 nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or a homologue, derivative or analogue thereof and the second primer molecule is preferably derived from the antisense strand of said gene.

Those skilled in the art will be aware that it is not essential to the performance of the
10 invention that the primer molecules be derived from the same gene.

According to this embodiment, the nucleic acid primer molecule may further consist of a combination of any of the nucleotides adenine, cytidine, guanine, thymidine, or inosine, or functional analogues or derivatives thereof, capable of being incorporated into
15 a polynucleotide molecule provided that it is capable of hybridising under at least low stringency conditions to the nucleic acid molecule set forth in SEQ ID NO:1 or SEQ ID NO:3 or a homologue, analogue or derivative thereof.

The nucleic acid primer molecules may further be each contained in an aqueous pool
20 comprising other nucleic acid primer molecules. More preferably, the nucleic acid primer molecule is in a substantially pure form.

The nucleic acid template molecule may be in a recombinant form, in a virus particle, bacteriophage particle, yeast cell, animal cell, or a plant cell. Preferably, the related
25 genetic sequence originates from a cell, tissue, or organ derived from an animal species or a human. More preferably, the related genetic sequence originates from a cell, tissue, or organ derived from a human.

Accordingly, a third aspect of the present invention extends to an isolated nucleic acid
30 molecule which is at least 40% identical to the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or to a complementary strand thereof or a homologue, analogue

or derivative thereof.

Preferably, the percentage identity to SEQ ID NO:1 or SEQ ID NO:3 is at least about 55%, still more preferably at least about 65%, yet still more preferably at least about 75-
5 80% and even still more preferably at least about 85-95%.

In an even more preferred embodiment, the present invention provides an isolated nucleic acid molecule which is at least 40% identical to the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 or to a complementary strand thereof or a homologue,
10 analogue or derivative thereof and is capable of hybridising under at least low stringency conditions to a nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3.

In a particularly preferred embodiment, the isolated nucleic acid molecule described herein is further capable of encoding a sequence of amino acids which is capable of
15 carrying out the enzyme reaction catalysed by a α -N-acetylglucosaminidase enzyme.

The isolated nucleic acid molecule of the present invention is also useful for developing a genetic construct comprising a sense molecule, for the expression or over-expression of α -N-acetylglucosaminidase in prokaryotic or eukaryotic cells. Particularly preferred
20 eukaryotic cells include CHO cells but may be in any other suitable mammalian cells or cell lines or non-mammalian cells such as yeast or insect cells.

The term "sense molecule" as used herein shall be taken to refer to an isolated nucleic acid molecule of the invention as described herein, which is provided in a format suitable
25 for its expression to produce a recombinant polypeptide, when said sense molecule is introduced into a host cell.

In a particularly preferred embodiment, a sense molecule which encodes the α -N-acetylglucosaminidase comprises a sequence of nucleotides set forth in SEQ ID NO:1 or
30 SEQ ID NO:3 or a complementary strand, homologue, analogue or derivative thereof.

- 20 -

In a most particularly preferred embodiment, the sense molecule of the invention comprises the sequence of nucleotides set forth in SEQ ID NO:1 or a complementary strand, homologue, analogue or derivative thereof.

5 Those skilled in the art will be aware that expression of a sense molecule may require the nucleic acid molecule of the invention to be placed in operable connection with a promoter sequence to produce a "sense construct". The choice of promoter for the present purpose may vary depending upon the level of expression of the sense molecule required and/or the tissue-specificity or developmental-specificity of expression of the
10 sense molecule which is required. The sense construct may further comprise a terminator sequence and be introduced into a suitable host cell where it is capable of being expressed to produce a recombinant polypeptide gene product.

In the context of the present invention, a sense molecule which corresponds to a genetic
15 sequence or isolated nucleic acid molecule which encodes α -N-acetylglucosaminidase polypeptide or a homologue, analogue or derivative thereof, placed operably under the control of a suitable promoter sequence, is introduced into a cell using any suitable method for the transformation of said cell and said genetic sequence or isolated nucleic acid molecule is expressed therein to produce said polypeptide.

20

The present invention clearly extends to genetic constructs designed to facilitate expression of any nucleic acid molecule described herein.

A genetic construct of the present invention comprises the foregoing sense molecule,
25 placed operably under the control of a promoter sequence capable of regulating the expression of the said nucleic acid molecule in a prokaryotic or eukaryotic cell, preferably a mammalian cell such as a CHO cell, a yeast cell, insect cell or bacterial cell. The said genetic construct optionally comprises, in addition to a promoter and sense molecule, a terminator sequence.

30

The term "terminator" refers to a DNA sequence at the end of a transcriptional unit

- 21 -

which signals termination of transcription. Terminators are 3'-non-translated DNA sequences containing a polyadenylation signal, which facilitates the addition of polyadenylate sequences to the 3'-end of a primary transcript. Terminators active in plant cells are known and described in the literature. They may be isolated from
5 bacteria, fungi, viruses, animals and/or plants.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box
10 sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. A promoter is usually, but not necessarily, positioned upstream or 5', of a structural gene, the expression of which it regulates. Furthermore, the regulatory elements comprising a promoter are usually
15 positioned within 2 kb of the start site of transcription of the gene.

In the present context, the term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of said sense molecule in a cell.

20

Preferred promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression of the sense molecule and/or to alter the spatial expression and/or temporal expression of said sense molecule. For example, regulatory elements which confer copper inducibility may be placed adjacent to a heterologous
25 promoter sequence driving expression of a sense molecule, thereby conferring copper inducibility on the expression of said molecule.

Placing a sense molecule under the regulatory control of a promoter sequence means positioning the said molecule such that expression is controlled by the promoter
30 sequence. Promoters are generally positioned 5' (upstream) to the genes that they control. In the construction of heterologous promoter/structural gene combinations it is

- 22 -

generally preferred to position the promoter at a distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting, i.e., the gene from which the promoter is derived. As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting, i.e., the genes from which it is derived. Again, as is known in the art, some variation in this distance can also occur.

10 Examples of promoters suitable for use in genetic constructs of the present invention include viral, fungal, bacterial, animal and plant derived promoters capable of functioning in animal, human, yeast, insect or bacterial cells. The promoter may regulate the expression of the said molecule constitutively, or differentially with respect to the tissue in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, or plant pathogens, or metal ions, amongst others. Preferably, the promoter is capable of regulating expression of a sense molecule in a cell derived from an animal species or human.

20 In a particularly preferred embodiment, the promoter is derived from the genomic gene encoding α -N-acetylglucosaminidase, preferably the human α -N-acetylglucosaminidase gene. In a more preferred embodiment, however, the promoter is derived from the nucleotide sequence set forth in SEQ ID NO:3 or is at least capable of hybridising to nucleotide residues 1 to 989 of SEQ ID NO:3 or at least 20 contiguous nucleotides derived therefrom.

In an even more particularly preferred embodiment, the promoter is the CMV promoter sequence or a promoter sequence derived therefrom.

30 An alternative embodiment of the invention is directed to a genetic construct comprising a promoter or functional derivative, part fragment, homologue, or analogue thereof,

- 23 -

derived from the α -N-acetylglucosaminidase genomic gene defined by SEQ ID NO: 3.

Preferably, said genetic construct further comprises the α -N-acetylglucosaminidase sequence defined by SEQ ID NO:1 placed in operably connection with said promoter.

5

A further aspect of the present invention is directed to synthetic α -N-acetylglucosaminidase or like molecule.

The term "synthetic" as used herein shall be taken to include both recombinant and
10 chemically-synthesised molecules produced by the sequential addition of amino acid residues or groups of amino acid residues in defined order.

In one embodiment, the invention relates to recombinant α -N-acetylglucosaminidase or like molecule encoded by or expressed from the nucleic acid molecules as hereinbefore
15 described.

In another embodiment, the synthetic α -N-acetylglucosaminidase or like molecule comprises a sequence of amino acids which is at least 40% identical to the amino acid sequence set forth in any one of SEQ ID Nos:2, 4, 5 or 6.

20

More preferably, the percentage identity is at least 60% and still more preferably at least 80% or 85-90%.

A particularly preferred embodiment of the present invention provides a synthetic α -N-acetylglucosaminidase as hereinbefore defined which comprises a sequence of amino acids
25 substantially as set forth in any one of SEQ ID Nos:2, 4, 5 or 6 or a homologue, analogue or derivative thereof.

For the purposes of nomenclature, the amino acid sequence set forth in SEQ ID NO:2
30 comprises the full-length translation product of the human α -N-acetylglucosaminidase gene (i.e. hereinafter referred to as the " α -N-acetylglucosaminidase polypeptide" or "SEQ ID

- 24 -

NO:2") produced by expression of either the cDNA sequence defined by SEQ ID NO:1 or the genomic gene defined by SEQ ID NO:3. The α -N-acetylglucosaminidase polypeptide comprises at least seven potentially-glycosylated Asn residues, at positions 261, 272, 435, 503, 513, 526 and 532. Furthermore, the amino acid sequence of the α -N-acetylglucosaminidase polypeptide may comprise a signal peptide of approximately 23 amino acid residues in length, with a probable site for signal peptide peptidase cleavage occurring between Gly₂₃ and Asp₂₄.

The amino acid sequences set forth in SEQ ID Nos:4-6 relate to N-terminal and internal (i.e. CNBr) amino acid sequences derived from human α -N-acetylglucosaminidase, purified as described in Example 1. As described in Example 2, the purified form of the enzyme comprises two polypeptides having approximate molecular weights of 82 and 77 kDa. The sequence set forth in SEQ ID NO:4 relates to the N-terminal sequence of the 82 kDa polypeptide, while SEQ ID NO:5 relates to the N-terminal sequence of the 77 kDa polypeptide. Furthermore, SEQ ID NO:4 comprises amino acids residues 24-43 of SEQ ID NO:2, while SEQ ID NO:5 comprises amino acid residues 59-76 of SEQ ID NO:2.

The amino acid sequence defined by SEQ ID NO:6 relates to the CNBr-cleaved peptide of purified human α -N-acetylglucosaminidase. This amino acid sequence aligns with amino acid residues 540-554 of the α -N-acetylglucosaminidase polypeptide (SEQ ID NO:2).

In the present context, "homologues" of a polypeptide refer to those polypeptides, enzymes or proteins which have a similar α -N-acetylglucosaminidase enzyme activity, notwithstanding any amino acid substitutions, additions or deletions thereto. A homologue may be isolated or derived from the same or another animal species.

Furthermore, the amino acids of a homologous polypeptide may be replaced by other amino acids having similar properties, for example hydrophobicity, hydrophilicity, hydrophobic moment, charge or antigenicity, and so on.

- 25 -

"Analogues" encompass α -N-acetylglucosaminidase polypeptides and peptide derivatives thereof notwithstanding the occurrence of any non-naturally occurring amino acid analogues therein.

5 The term "derivative" in relation to the polypeptides of the invention refer to mutants, parts or fragments of a functional molecule. Derivatives include modified peptides in which ligands are attached to one or more of the amino acid residues contained therein, such as carbohydrates, enzymes, proteins, polypeptides or reporter molecules such as radionuclides or fluorescent compounds. Glycosylated, fluorescent, acylated or alkylated
10 forms of the subject peptides are particularly contemplated by the present invention. Additionally, derivatives of a polypeptide may comprise fragments or parts of an amino acid sequence disclosed herein and are within the scope of the invention, as are homopolymers or heteropolymers comprising two or more copies of the subject polypeptides. Procedures for derivatizing peptides are well-known in the art.

15

Accordingly, this aspect of the present invention is directed to any proteinaceous molecule comprising an amino acid sequence corresponding to the full length mammalian α -N-acetylglucosaminidase enzyme or to a like molecule. The like molecule, therefore, comprises parts, derivatives and/or portions of the α -N-acetylglucosaminidase enzyme
20 whether functional or not.

Preferably, the mammal is human but may be of non-human origin as contemplated above.

The synthetic or recombinant α -N-acetylglucosaminidase of the present invention may
25 comprise an amino acid sequence corresponding to the naturally occurring amino acid sequence or may contain single or multiple amino acid substitutions, deletions and/or additions. The length of the amino acid sequence may range from a few residues to a full length molecule.

30 Amino acid substitutions are typically of single residues. Amino acid insertions will usually be in the order of about 1-10 amino acid residues and deletions will range from

- 26 -

about 1-20 residues. Preferably, deletions or insertions are made in adjacent pairs, i.e. a deletion of two residues or insertion of two residues.

Amino acid insertional derivatives of α -N-acetylglucosaminidase of the present invention include amino and/or carboxyl terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Insertional amino acid sequence variants are those in which one or more amino acid residues are introduced into a predetermined site in the protein although random insertion is also possible with suitable screening of the resulting product. Deletional variants are characterised by the removal of one or more amino acids from the sequence. Substitutional amino acid variants are those in which at least one residue in the sequence has been removed and a different residue inserted in its place. Typical substitutions are those made in accordance with the following Table 2:

The amino acid variants referred to above may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis (Merrifield synthesis) and the like, or by recombinant DNA manipulations. Techniques for making substitution mutations at predetermined sites in DNA having known or partially known sequence are well known and include, for example, M13 mutagenesis. The manipulation of DNA sequence to produce variant proteins which manifest as substitutional, insertional or deletional variants are conveniently elsewhere described such as Sambrook *et al*, 1989 *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

The derivatives or like molecules include single or multiple substitutions, deletions and/or additions of any component(s) naturally or artificially associated with the α -N-acetylglucosaminidase enzyme such as carbohydrate, lipid and/or other proteinaceous moieties. For example, the present invention extends to glycosylated and non-glycosylated forms of the molecule. All such molecules are encompassed by the expression "mutants", "derivatives", "fragments", "portions" and "like" molecules. These molecules may be active or non-active and may contain specific regions, such as a catalytic region. Particularly, preferred derivative molecules include those with altered glycosylation

- 27 -

patterns relative to the naturally occurring molecule. Even more particularly, the recombinant molecule is more highly glycosylated than the naturally occurring molecule. Such highly glycosylated derivatives may have improved take-up properties and enhanced half-lives.

5

As indicated in the Examples, the molecular weight of purified human α -N-acetylglucosaminidase (i.e. 82kDa and 77kDa) and recombinant mammalian α -N-acetylglucosaminidase produced in CHO cells (i.e. 89 kDa and 79 kDa) are greater than the deduced molecular weight of the α -N-acetylglucosaminidase polypeptide set forth in
10 SEQ ID No:2 (i.e. 70 kDa), suggesting that the purified and recombinant polypeptide are post-translationally modified. The data presented in Example 8 indicate further that the recombinant α -N-acetylglucosaminidase enzyme produced in CHO cells, at least, is glycosylated and that the difference in molecular weight determined for the recombinant polypeptides and the polypeptide of SEQ ID No: 2 is due almost entirely to glycosylation
15 of the recombinant polypeptide by CHO cells. As shown in Example 9, the glycosylated recombinant α -N-acetylglucosaminidase polypeptide exhibits enzymatic activity.

The present invention also extends to synthetic α -N-acetylglucosaminidase or like molecules when fused to other proteinaceous molecules. The latter may include another
20 enzyme, reporter molecule, purification site or an amino acid sequence which facilitates transport of the molecule out of a cell, such as a signal sequence.

The present invention extends further to post-translational modifications to the α -N-acetylglucosaminidase enzyme. The modifications may be made to the naturally occurring
25 enzyme or following synthesis by recombinant techniques. The modifications may be at the structural level or at, for example, the electrochemical level such as modifying net charge or structural conformation of the enzyme.

Such modification may be important to facilitate entry or penetration of the enzyme into
30 selected tissues such as cartilage or blood brain barriers or to increase circulation half-life.

- 28 -

Analogues of α -N-acetylglucosaminidase contemplated herein include, but are not limited to, modifications to side chains, incorporation of unnatural amino acids and/or their derivatives during peptide synthesis and the use of crosslinkers and other methods which impose conformational constraints on the enzyme.

5

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with NaBH_4 ; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; 10 trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5'-phosphate followed by reduction with NaBH_4 .

The guanidino group of arginine residues may be modified by the formation of 15 heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivatisation, for example, to a corresponding amide.

20

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4- 25 chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuric-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N- 30 bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with

- 29 -

tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with
5 diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline,
10 phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. Non-naturally occurring amino acids contemplated by the present invention are incorporated herein, as Table 3.

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-
15 bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with $n=1$ to $n=6$, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, the enzyme could be conformationally constrained by,
20 for example, incorporation of C_α and N_α -methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

25 Electrochemical modifications of α -N-acetylglucosaminidase include interaction with polylysine or polyethylene glycol or other agent which effects an overall change to the net charge of the enzyme.

Advantageously, the recombinant α -N-acetylglucosaminidase is a biologically pure
30 preparation meaning that it has undergone some purification away from other proteins and/or non-proteinaceous material. The purity of the preparation may be represented as at least

- 30 -

40% of the enzyme, preferably at least 60%, more preferably at least 75%, even more preferably at least 85% and still more preferably at least 95% relative to non- α -N-acetylglucosaminidase material as determined by weight, activity, amino acid homology or similarity, antibody reactivity or other convenient means.

5

Particularly preferred methods for the preparation and purification of recombinant α -N-acetylglucosaminidase are provided in Examples 7 and 8.

Those skilled in the art will be aware of the means of purifying a synthetic or recombinant
10 α -N-acetylglucosaminidase from several sources without undue experimentation and for expressing the degree of purity of such a purified preparation of the enzyme.

The present invention further contemplates antibodies to α -N-acetylglucosaminidase and preferably synthetic α -N-acetylglucosaminidase or like molecule. The antibodies may be
15 polyclonal or monoclonal, naturally occurring or synthetic (including recombinant, fragment or fusion forms). Such antibodies will be useful in developing immunoassays for α -N-acetylglucosaminidase and for identifying additional genetic sequences which are capable of expressing α -N-acetylglucosaminidase polypeptides or homologues, analogues or derivatives thereof.

20

Both polyclonal and monoclonal antibodies are obtainable by immunisation with an appropriate synthetic or recombinant gene product, or epitope, or peptide fragment of a gene product, in particular a α -N-acetylglucosaminidase polypeptide or a homologue, analogue or derivative thereof.

25

Alternatively, fragments of antibodies may be used, such as Fab fragments. The present invention extends further to encompass recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

30

A further aspect of the present invention contemplates a method of screening for mutations

or other aberrations in the α -N-acetylglucosaminidase gene in a human or animal patient. Such a method may be accomplished in a number of ways including isolating a source of DNA to be tested or mRNA therefrom and hybridising thereto a nucleic acid molecule as hereinbefore described. Generally, the nucleic acid is probe or primer size and polymerase

5 chain reaction is a convenient means by which to analyse the RNA or DNA. Other suitable assays include the ligation chain reaction and the strand displacement amplification methods. The α -N-acetylglucosaminidase sequence can also be determined and compared to the naturally occurring sequence. Such methods may be useful in adults and children and may be adapted for a pre-natal test. The DNA to be tested includes a genomic sample

10 carrying the α -N-acetylglucosaminidase gene, a cDNA clone and/or amplification product.

In accordance with this aspect of the present invention there is provided a method for screening for aberrations in the α -N-acetylglucosaminidase gene including the absence of such a gene or a portion or a substantial portion thereof comprising isolating a sample of

15 DNA or mRNA corresponding to a region of said DNA and contacting same with an oligonucleotide probe capable of hybridising to one or more complementary sequences within the α -N-acetylglucosaminidase gene and then detecting the hybridisation, the extent of hybridisation or the absence of hybridisation.

20 Alternatively, the probe is a primer and capable of directing amplification of one or more regions of said α -N-acetylglucosaminidase gene and the amplification products and/or profile of amplification products is compared to an individual carrying the full gene or to a reference data base.

25 Conveniently, the amplification products are sequenced to determine the presence or absence of the full gene.

The present invention extends to the use of any and all DNA-based or nucleic acid-based hybridisation and/or polymerase chain reaction formats as described herein, for the

30 diagnosis of a disorder involving the α -N-acetylglucosaminidase gene in a human or animal patient.

The present invention further extends to a method of treating patients suffering from α -N-acetylglucosaminidase deficiency, such as in MPS-IIIB, said method comprising administering to said patient an effective amount of α -N-acetylglucosaminidase or active like form thereof.

5

Preferably, the α -N-acetylglucosaminidase is in recombinant form. Such a method is referred to as "enzyme therapy". Alternatively, gene therapy can be employed including introducing an active gene (i.e. a nucleic acid molecule as hereinbefore described) or to parts of the gene or other sequences which facilitate expression of a naturally occurring α -

10 N-acetylglucosaminidase gene.

Administration of α -N-acetylglucosaminidase for enzyme therapy may be by oral, intravenous, suppository, intraperitoneal, intramuscular, intranasal, intradermal or subcutaneous administration or by infusion or implantation. The α -N-
15 acetylglucosaminidase is preferably as hereinbefore described including active mutants or derivatives thereof and glycosylation variants thereof. Administration may also be by way of gene therapy including expression of the gene by inclusion of the gene in viral vectors which are introduced into the animal (e.g. human) host to be treated. Alternatively, the gene may be expressed in a bacterial host which is then introduced and becomes part of the
20 bacterial flora in the animal to be tested.

Still yet another aspect of the present invention is directed to a pharmaceutical composition comprising synthetic (e.g. recombinant) α -N-acetylglucosaminidase or like molecule, including active derivatives and fragments thereof, alone or in combination with other
25 active molecules. Such other molecules may act synergistically with the enzyme or facilitates its entry to a target cell. The composition will also contain one or more pharmaceutically acceptable carriers and/or diluents. The composition may alternatively comprise a genetic component useful in gene therapy.

30 The active ingredients of the pharmaceutical composition comprising the synthetic or recombinant α -N-acetylglucosaminidase or mutants or fragments or derivatives thereof are

contemplated to exhibit excellent activity in treating patients with a deficiency in the enzyme when administered in an amount which depends on the particular case. The variation depends, for example, on the patient and the α -N-acetylglucosaminidase used. For example, from about 0.5 ug to about 20 mg of enzyme per animal body or, depending on the animal and other factors, per kilogram of body weight may be administered. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily, weekly, monthly or in other suitable time intervals or the dose may be proportionally reduced as indicated by the exigencies of the situation. Accordingly, alternative dosages in the order of 1.0 μ g to 15 mg, 2.0 μ g to 10 mg or 10 μ g to 5mg may be administered in a single or as part of multiple doses. The active compound may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intramuscular, subcutaneous, intranasal, intradermal or suppository routes or implanting (eg using slow release molecules). Depending on the route of administration, the active ingredients which comprise a synthetic (e.g. recombinant) α -N-acetylglucosaminidase or fragments, derivatives or mutants thereof may be required to be coated in a material to protect same from the action of enzymes, acids and other natural conditions which may inactivate said ingredients. For example, the low lipophilicity of α -N-acetylglucosaminidase will allow it to be destroyed in the gastrointestinal tract by enzymes capable of cleaving peptide bonds and in the stomach by acid hydrolysis. In order to administer the vaccine by other than parenteral administration, the enzyme will be coated by, or administered with, a material to prevent its inactivation. For example, the enzyme may be administered in an adjuvant, co-administered with enzyme inhibitors or in liposomes. Adjuvant is used in its broadest sense and includes any immune stimulating compound such as interferon. Adjuvants contemplated herein include resorcinols, non-ionic surfactants such as polyoxyethylene oleyl ether and n-hexadecyl polyethylene ether. Conveniently, the adjuvant is Freund's Complete or Incomplete Adjuvant. Enzyme inhibitors include pancreatic trypsin inhibitor, diisopropylfluorophosphate (DEP) and trasylol. Liposomes include water-in-oil-in-water emulsions as well as conventional liposomes.

30

The active compound may also be administered in dispersions prepared in glycerol, liquid

polyethylene glycols, and/or mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

- 5 The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating
- 10 action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of
- 15 dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the
- 20 compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

- Sterile injectable solutions are prepared by incorporating the active compound in the required amount in the appropriate solvent with various of the other ingredients
- 25 enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredient(s) into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the
- 30 freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

- 35 -

- When the α -N-acetylglucosaminidase of the present invention is suitably protected as described above, the composition may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly
- 5 with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be
- 10 between about 5 to about 80% of the weight of the unit. The amount of active compound in the vaccine compositions is such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared, so that an oral dosage unit form contains between about 0.5 ug and 20 mg of active compound.
- 15 The tablets, troches, pills, capsules and the like may also contain the following: a binder such as gum tragacanth, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavoring agent such as peppermint, oil of wintergreen, or
- 20 cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as
- 25 preservatives, a dye and flavoring such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound may be incorporated into sustained-release preparations and formulations.
- 30 As used herein "pharmaceutically acceptable carriers and/or diluents" include any and all solvents, dispersion media, aqueous solutions, coatings, antibacterial and antifungal agents,

- 36 -

isotonic and absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the pharmaceutical compositions is contemplated. Supplementary active ingredients can also
5 be incorporated into the compositions.

The present invention further relates to the use of α -N-acetylglucosaminidase or active fragment, mutant or derivative thereof in the manufacture of a medicament for the treatment of patients suffering from a deficiency in the naturally occurring enzyme (e.g.
10 MPS-IIIB).

The present invention is further described with reference to the following non-limiting Examples.

15

EXAMPLE 1

Purification of α -N-Acetylglucosaminidase

α -N-acetylglucosaminidase was purified according to the method described in Weber *et al.* (1996). Enzyme was purified to homogeneity from human placenta. Evidence of
20 purity is shown following SDS/PAGE which is represented in Figure 1. All samples were reduced with dithiothreitol prior to electrophoresis.

EXAMPLE 2

Characterisation of α -N-Acetylglucosaminidase

25

Results presented in Figure 1 show two polypeptides of about 82kDa and 77 kDa molecular weight, which correspond to α -N-acetylglucosaminidase polypeptides purified from human placenta according to Example 1.

30

EXAMPLE 3

Amin Acid Sequence Determination

The N-terminal amino acid sequences the 77 kDa and 82 kDa α -N-acetylglucosaminidase polypeptides, in addition to the amino acid sequence of an internal CNBr cleavage product of these peptides, were determined using the methods of Weber *et al.* (1996).

5

The amino acid sequences are shown in Table 4.

- 38 -

TABLE 4

N-Terminal amino acid sequences (SEQ ID NO:4 and SEQ ID No:5) and
CNBr peptide sequence (SEQ ID No:6) determined from purified human
 α -N-Acetylglucosaminidase

5

polypeptide 82 kDa	DEAREAAAVRALVARLLGPG
polypeptide 77 kDa	KPGLDTYSLGGGGAAX ¹ VR
CNBr peptide	WRLLLTSAPSLX ¹ TX ¹ P

10

X¹ no residue could be identified for this position, indicating that this residue could be
phosphorylated or glycosylated.

15

EXAMPLE 4

Cloning of α -N-Acetylglucosaminidase cDNA

20

Oligonucleotide probes were prepared based on the partial amino acid sequences obtained
for the purified α -N-acetylglucosaminidase polypeptides (Example 3). The probes were
subsequently used to screen a human peripheral blood leukocyte cDNA library. An
approximately 2.6 kbp cDNA clone was isolated encoding most of the sequence of human
 α -N-acetylglucosaminidase (SEQ ID NO:1).

25

The remaining α -N-acetylglucosaminidase coding sequence was obtained from the
nucleotide sequence of the corresponding genomic gene (SEQ ID NO:3), isolated by
hybridisation to a human chromosome 17 library (Weber *et.al.* 1996).

30

- 39 -

The complete open reading frame is 2232 nucleotides long and encodes a 743 (plus stop codon) amino acid protein. The predicted molecular mass of the longest mature protein (minus the 23 amino acid N-terminal signal peptide) is about 79,622 daltons.

- 5 The amino acid sequence of α -N-acetylglucosaminidase is shown in SEQ ID NO:2. The deduced molecular weight of the desired amino acid sequence of α -N-acetylglucosaminidase is approximately 70kDa. The probable site of signal peptide peptidase cleavage is between amino acids 23 and 24. There are seven potential N-glycosylation sites in the sequence.

10

The nucleotide sequence of the corresponding α -N-acetylglucosaminidase genomic gene (SEQ ID No:3) comprises 10380 bp including 889 bp of 5' upstream sequence corresponding to at least at part of the α -N-acetylglucosaminidase promoter sequence, in addition to the nucleotide sequences of introns I, II, II, IV, V, in addition to 1326 bp of 3'-

15 untranslated sequence.

EXAMPLE 5

Construction of an expression vector comprising the α -N-Acetylglucosaminidase cDNA sequence

20

- The cDNA insert of λ clone pbl 33, containing bases 107 to 2575 of the α -N-acetylglucosaminidase cDNA was excised with *EcoRI* and subcloned into pBluescript II SK-(Stratagene). A 178 bp *XmaI* fragment (bases 1 to 178 of the α -N-acetylglucosaminidase cDNA) from cosmid sub-clone 6.3, containing the start codon, was
- 25 cloned into the pBluescript subclone to produce a full-length cDNA sequence in addition to 101 bp of 5' non-translated sequence as well as 245 bp of 3' non-translated region including the polyadenylation-site, the polyA-tail and linkerDNA. The full length cDNA was directionally cloned into the pCDNA3 expressionvector (Invitrogen) via the *EcoRI* and *BamHI* sites.

30

EXAMPLE 6

- 40 -

Expressi n of rec mbinant α -N-acetylglucosaminidase

- Chinese Hamster Ovary (CHO) cells were transfected with expressionvector using the DOTAP transfection reagent (Boehringer Mannheim) according to the manufacturers
- 5 instructions. Cells were grown in Ham's F12 medium, 10% (v/v) fetal calf serum, penicillin and streptomycin sulfate at 100 μ g/ml each. Cells were grown in nonselective medium for 48 h and then incubated in medium containing 750 μ g/ml G418 sulfate (Genitacin) until resistant colonies emerged.
- 10 Single cell clones were grown up and 26 of them were tested for expression of recombinant α -N-acetylglucosaminidase with a fluorogenic α -N-acetylglucosaminidase substrate. (i.e. N-acetylglucosamine α -linked to 4-methylumbelliferone)

EXAMPLE 7

15 Large scale α -N-acetylglucosaminidase production

- 2 g of Cytodex 2 microcarrier beads were swollen in 250 ml of PBS for 3 h at 37°C with three changes of PBS and then autoclaved for 15 min at 120°C (wet cycle). The beads were then rinsed with sterile growth medium (Coons/DMEM, 10% v/v fetal calf serum,
- 20 penicillin and streptomycin sulfate at 100 μ g/ml each and 0.1% w/v Pluronic F68) and transferred into a Techne stirrer culture flask. The microcarrier beads were inoculated with seven confluent 175 flasks of the cell clone showing the highest expression of recombinant α -N-acetylglucosaminidase. Growth medium was added up to 200 ml and the culture incubated with a stirrer speed of 20 rpm to achieve an even distribution of cells on the
- 25 beads. The cells were allowed to attach to the beads for 16 h at low speed then medium was added up to 500 ml and the stirrer speed increased to 30 rpm. After a growth phase of 48 to 72 h with daily aerating to allow gas exchange the beads were completely covered with cells and the medium was exchanged for production medium (Coons/DMEM, no fetal calf serum, penicillin and streptomycin sulfate at 100 μ g/ml each, 0.1% w/v Pluronic F68
- 30 and 5 mM NH_4Cl). The glucose concentration was monitored daily and the medium replaced, when glucose fell below 5 mM every 203 days. The harvested medium contained

approximately 2 mg α -N-acetylglucosaminidase protein per dm³ of production medium.

EXAMPLE 8

Purification of recombinant α -N-acetylglucosaminidase

5

Production medium was dialysed against 50 mM NaAc pH 5.5 and loaded onto a heparin-agarose column equilibrated in the same buffer. After washing with NaAc buffer and NaAc/50 mM NaCl the column was eluted with 75 mM NaCl in NaAc buffer. The eluate was dialysed against 20 mM Tris/HCl pH 7.5, loaded onto a DEAE Sephacel column,
10 washed with 25 mM NaCl in 20 mM Tris/HCl and then eluted with 50 and 75 mM NaCl in 20 mM Tris/HCl respectively.

SDS-PAGE of the two eluates showed two polypeptide bands associated with enzyme activity with apparent molecular weights of 79 and 89 kDa. The smaller α -N-
15 acetylglucosaminidase was eluted predominantly in the 50 mM NaCl fraction whereas the 89 kDa α -N-acetylglucosaminidase polypeptide was enriched in the 75 mM NaCl fraction (Fig. 2).

The difference in apparent molecular weight of the recombinant α -N-
20 acetylglucosaminidase polypeptides is due to the presence of additional carbohydrate side chains, since a digest with PNGase F, which cleaves off N-glycosylation groups, reduced both the 79 kDa and 89 kDa polypeptides to the polypeptide band having an apparent molecular weight of about 70 kDa (Fig. 2), which corresponds to the approximate molecular weight deduced from primary amino acid sequence data (SEQ ID No:2).

25

EXAMPLE 9

Characteristics of recombinant α -N-acetylglucosaminidase

No differences were observed between the enzyme activities of the 79 and 89 kDa
30 recombinant α -N-acetylglucosaminidase polypeptides produced in CHO cells according to Example 7 and 8. With the fluorogenic N-acetylglucosamine α -linked to 4-

- 42 -

methyumbelliferone (4-MU) substrate, the enzyme has a pH-optimum of 4.6 with a k_M of 5.34 mM and a V_{max} of 3.97×10^6 pmol/min/mg. Towards a ^3H -labelled disaccharide substrate it should have a pH-optimum of 4.1 with a k_M of 0.0166 mM and a V_{max} of 4.48×10^4 pmol/min/mg.

5

EXAMPLE 10

Mutational analysis of Sanfilippo B patients

Genomic DNA is isolated from cultivated skin fibroblasts of patients by extraction with
10 Phenol/Chloroform and used to amplify the eight exons and adjacent intronic sequences individually by PCR.

Primer sequences used in the amplification reaction are readily determined from the nucleotide sequences of the α -N-acetylglucosaminidase cDNA and genomic clones set
15 forth in SEQ ID No:1 or SEQ ID No:3. Amplification conditions are also readily determined without undue experimentation. Procedures for the design of PCR primers and amplification conditions are described in detail, for example, by McPherson *et al.* (1991). Differences in the primary sequence can be identified by separating the PCR products on a polyacrylamide gel under non-denaturing conditions (SSCP gels). Base changes,
20 insertions and deletions will lead to a different band pattern compared with the wildtype in most of the cases, which can be visualised either by autoradiography of the gel after labelling the DNA during the PCR or by staining unlabelled DNA in the gel with silver. PCR products which show a different band pattern are sequences to identify the change. Other patient samples can be tested for mutations and polymorphism that were found by
25 hybridisation with wildtype- and mutation-specific oligonucleotides (ASO).

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also
30 includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or

- 43 -

more of said steps or features.

REFERENCES:

1. Ausubel, F. M., Brent, R., Kingston, RE, Moore, D.D., Seidman, J.G., Smith, J.A., and
5 Struhl, K. (1987). *In: Current Protocols in Molecular Biology*. Wiley Interscience (ISBN
047150338).
2. Hopwood JJ (1989) *In: "Heparin: Chemical and Biological Properties, Clinical
Applications"* (Lane DW and Lindahl U, eds.), 190-229, Edward Arnold, London.
- 10 3. McKusick V and Neufeld E (1983) *In: "The Metabolic Basis of Inherited Disease"*
(Stanbury JB, Wyngaarden JB, Fredrickson DS, Goldstein JL and Brown MS, eds), 5th
Ed., 751-771, McGraw-Hill, New York.
- 15 4. McPherson, M.J., Quirke, P. and Taylor, G.R., (1991) *In: PCR A Practical Approach*.
Oxford University Press, Oxford. (ISBN 0-19-96322L-X).
5. O'Brien JS, (1972) *Proc. Natl. Acad. Sci. USA* 69: 1720-1722.
- 20 6. Sambrook, J., Fritsch, E., and Maniatis, T. (1989) *In: "Molecular Cloning" a
laboratory manual*, Cold Spring Harbour.
7. Von Figura, K, and Kresse H, (1972) *Biochem Biophys. Res. Commun*, 48: 262-269
- 25 8. Weber B, Scott H, Blanch L, Clements P, Morris CP, Anson D, Hopwood J, (1996)
Nature Genetics (submitted)

- 45 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: WOMEN'S AND CHILDREN'S HOSPITAL

(US ONLY:HOPWOOD JJ, WEBER B SCOTT HS, BLANCH LC, ANSON D)

(ii) TITLE OF INVENTION: SYNTHETIC MAMMALIAN α -N-ACETYLGLUCOSAMINIDASE AND
GENETIC SEQUENCES ENCODING SAME

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: AU PN6748/95
(B) FILING DATE: 23-NOV-1995

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:PCT INTERNATIONAL
(B) FILING DATE: 22-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: HUGHES, DR E JOHN L
(C) REFERENCE/DOCKET NUMBER: MRO/EJH/JMC

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE:+61 3 9254 2777
(B) TELEFAX: +61 3 9254 2770

- 46 -

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2575 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens
- (F) TISSUE TYPE: Peripheral Blood
- (G) CELL TYPE: Leukocyte

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 102..2333

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

CCCCGGGCTTA GCCTTCGGGT CCACGTGGCC GGAGGCCGGC AGCTGATTGG ACGCGGGCCG      60
CCCCACCCCC TGGCCGTCGC GGGACCCGCA GGACTGAGAC C ATG GAG GCG GTG      113
                                         Met Glu Ala Val
                                         1

GCG GTG GCC GCG GCG GTG GGG GTC CTT CTC CTG GCC GGG GCC GGG GGC      161
Ala Val Ala Ala Ala Val Gly Val Leu Leu Ala Gly Ala Gly Gly
   5              10              15              20

GCG GCA GGC GAC GAG GCC CGG GAG GCG GCG GCC GTG CGG GCG CTC GTG      209
Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val
              25              30              35

GCC CGG CTG CTG GGG CCA GGC CCC GCG GCC GAC TTC TCC GTG TCG GTG      257
Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe Ser Val Ser Val
              40              45              50

GAG CGC GCT CTG GCT GCC AAG CCG GGC TTG GAC ACC TAC AGC CTG GGC      305
Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly
              55              60              65

GGC GGC GGC GCG GCG CGC GTG CGG GTG CGC GGC TCC ACG GGC GTG GCG      353
Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser Thr Gly Val Ala
              70              75              80

GCC GCC GCG GGG CTG CAC CGC TAC CTG CGC GAC TTC TGT GGC TGC CAC      401
Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe Cys Gly Cys His
              85              90              95              100

```

- 47 -

GTG GCC TGG TCC GGC TCT CAG CTG CGC CTG CCG CGG CCA CTG CCA GCC	449
Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg Pro Leu Pro Ala	
105 110 115	
GTG CCG GGG GAG CTG ACC GAG GCC ACG CCC AAC AGG TAC CGC TAT TAC	497
Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg Tyr Arg Tyr Tyr	
120 125 130	
CAG AAT GTG TGC ACG CAA AGC TAC TCC TTC GTG TGG TGG GAC TGG GCC	545
Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp Trp Asp Trp Ala	
135 140 145	
CGC TGG GAG CGA GAG ATA GAC TGG ATG GCG CTG AAT GGC ATC AAC CTG	593
Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn Gly Ile Asn Leu	
150 155 160	
GCA CTG GCC TGG AGC GGC CAG GAG GCC ATC TGG CAG CGG GTG TAC CTG	641
Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln Arg Val Tyr Leu	
165 170 175 180	
GCC TTG GGC CTG ACC CAG GCA GAG ATC AAT GAG TTC TTT ACT GGT CCT	689
Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe Phe Thr Gly Pro	
185 190 195	
GCC TTC CTG GCC TGG GGG CGA ATG GGC AAC CTG CAC ACC TGG GAT GGC	737
Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His Thr Trp Asp Gly	
200 205 210	
CCC CTG CCC CCC TCC TGG CAC ATC AAG CAG CTT TAC CTG CAG CAC CGG	785
Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr Leu Gln His Arg	
215 220 225	
GTC CTG GAC CAG ATG CGC TCC TTC GGC ATG ACC CCA GTG CTG CCT GCA	833
Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro Val Leu Pro Ala	
230 235 240	
TTC GCG GGG CAT GTT CCC GAG GCT GTC ACC AGG GTG TTC CCT CAG GTC	881
Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val Phe Pro Gln Val	
245 250 255 260	
AAT GTC ACG AAG ATG GGC AGT TGG GGC CAC TTT AAC TGT TCC TAC TCC	929
Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn Cys Ser Tyr Ser	
265 270 275	
TGC TCC TTC CTT CTG GCT CCG GAA GAC CCC ATA TTC CCC ATC ATC GGG	977
Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe Pro Ile Ile Gly	
280 285 290	
AGC CTC TTC CTG CGA GAG CTG ATC AAA GAG TTT GGC ACA GAC CAC ATC	1025
Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly Thr Asp His Ile	
295 300 305	

- 48 -

TAT GGG GCC GAC ACT TTC AAT GAG ATG CAG CCA CCT TCC TCA GAG CCC Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro Ser Ser Glu Pro 310 315 320	1073
TCC TAC CTT GCC GCA GCC ACC ACT GCC GTC TAT GAG GCC ATG ACT GCA Ser Tyr Leu Ala Ala Ala Thr Thr Ala Val Tyr Glu Ala Met Thr Ala 325 330 335 340	1121
GTG GAT ACT GAG GCT GTG TGG CTG CTC CAA GGC TGG CTC TTC CAG CAC Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp Leu Phe Gln His 345 350 355	1169
CAG CCG CAG TTC TGG GGG CCC GCC CAG ATC AGG GCT GTG CTG GGA GCT Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala Val Leu Gly Ala 360 365 370	1217
GTG CCC CGT GGC CGC CTC CTG GTT CTG GAC CTG TTT GCT GAG AGC CAG Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe Ala Glu Ser Gln 375 380 385	1265
CCT GTG TAT ACC CGC ACT GCC TCC TTC CAG GGC CAG CCC TTC ATC TGG Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gly Gln Pro Phe Ile Trp 390 395 400	1313
TGC ATG CTG CAC AAC TTT GGG GGA AAC CAT GGT CTT TTT GGA GCC CTA Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu Phe Gly Ala Leu 405 410 415 420	1361
GAG GCT GTG AAC GGA GGC CCA GAA GCT GCC CGC CTC TTC CCC AAC TCC Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu Phe Pro Asn Ser 425 430 435	1409
ACC ATG GTA GGC ACG GGC ATG GCC CCC GAG GGC ATC AGC CAG AAC GAA Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile Ser Gln Asn Glu 440 445 450	1457
GTG GTC TAT TCC CTC ATG GCT GAG CTG GGC TGG CGA AAG GAC CCA GTG Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg Lys Asp Pro Val 455 460 465	1505
CCA GAT TTG GCA GCC TGG GTG ACC AGC TTT GCC GCC CGG CGG TAT GGG Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala Arg Arg Tyr Gly 470 475 480	1553
GTC TCC CAC CCG GAC GCA GGG GCA GCG TGG AGG CTA CTG CTC CGG AGT Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu Leu Leu Arg Ser 485 490 495 500	1601
GTG TAC AAC TGC TCC GGG GAG GCC TGC AGG GGC CAC AAT CGT AGC CCG Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His Asn Arg Ser Pro 505 510 515	1649

- 49 -

CTG GTC AGG CGG CCG TCC CTA CAG ATG AAT ACC AGC ATC TGG TAC AAC	1697
Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser Ile Trp Tyr Asn	
520 525 530	
CGA TCT GAT GTG TTT GAG GCC TGG CGG CTG CTG CTC ACA TCT GCT CCC	1745
Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu Thr Ser Ala Pro	
535 540 545	
TCC CTG GCC ACC AGC CCC GCC TTC CGC TAC GAC CTG CTG GAC CTC ACT	1793
Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu Leu Asp Leu Thr	
550 555 560	
CGG CAG GCA GTG CAG GAG CTG GTC AGC TTG TAC TAT GAG GAG GCA AGA	1841
Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr Glu Glu Ala Arg	
565 570 575 580	
AGC GCC TAC CTG AGC AAG GAG CTG GCC TCC CTG TTG AGG GCT GGA GGC	1889
Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu Arg Ala Gly Gly	
585 590 595	
GTC CTG GCC TAT GAG CTG CTG CCG GCA CTG GAC GAG GTG CTG GCT AGT	1937
Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu Val Leu Ala Ser	
600 605 610	
GAC AGC CGC TTC TTG CTG GGC AGC TGG CTA GAG CAG GCC CGA GCA GCG	1985
Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln Ala Arg Ala Ala	
615 620 625	
GCA GTC AGT GAG GCC GAG GCC GAT TTC TAC GAG CAG AAC AGC CGC TAC	2033
Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln Asn Ser Arg Tyr	
630 635 640	
CAG CTG ACC TTG TGG GGG CCA GAA GGC AAC ATC CTG GAC TAT GCC AAC	2081
Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu Asp Tyr Ala Asn	
645 650 655 660	
AAG CAG CTG GCG GGG TTG GTG GCC AAC TAC TAC ACC CCT CGC TGG CGG	2129
Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr Pro Arg Trp Arg	
665 670 675	
CTT TTC CTG GAG GCG CTG GTT GAC AGT GTG GCC CAG GGC ATC CCT TTC	2177
Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln Gly Ile Pro Phe	
680 685 690	
CAA CAG CAC CAG TTT GAC AAA AAT GTC TTC CAA CTG GAG CAG GCC TTC	2225
Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu Glu Gln Ala Phe	
695 700 705	
GTT CTC AGC AAG CAG AGG TAC CCC AGC CAG CCG CGA GGA GAC ACT GTG	2273
Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg Gly Asp Thr Val	
710 715 720	

- 50 -

GAC CTG GCC AAG AAG ATC TTC CTC AAA TAT TAC CCC GGC TGG GTG GCC 2321
Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro Gly Trp Val Ala
725 730 735 740

GGC TCT TGG TGATAGATTC GCCACCACTG GGCCTTGTTT TCCGCTAATT 2370
Gly Ser Trp

CCAGGGCAGA TTCCAGGGCC CAGAGCTGGA CAGACATCAC AGGATAACCC AGGCCTGGGA 2430

GGAGGCCCA CGGCCTGCTG GTGGGGTCTG ACCTGGGGGG ATTGGAGGGA AATGACCTGC 2490

CCTCCACCAC CACCCAAAGT GTGGGATTAA AGTACTGTTT TCTTTCCTACT TAAAAAAAAA 2550

AAAAAAGTCG AGCGGCCGCG AATTC 2575

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 743 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

(ix) **FEATURE:**

(A) NAME/KEY: Potentially-glycosylated Asn site,

(B) LOCATION: 261

(ix) FEATURE:

(A) NAME/KEY: Potentially-glycosylated Asn site,

(B) LOCATION: 272

(ix) **FEATURE:**

(A) NAME/KEY: Potentially-glycosylated Asn site.

(B) LOCATION: 435

(ix) FEATURE:

(A) NAME/KEY: Potentially-glycosylated Asn site.

(B) LOCATION: 503

(ix) FEATURE:

(A) NAME/KEY: Potentially-glycosylated Asn site.

(B) LOCATION: 513

(ix) **FEATURE:**

- 51 -

(A) NAME/KEY: Potentially-glycosylated Asn site,
 (B) LOCATION: 526

(ix) FEATURE:

(A) NAME/KEY: Potentially-glycosylated Asn site,
 (B) LOCATION: 532

Met Glu Ala Val Ala Val Ala Ala Ala Val Gly Val Leu Leu Leu Ala
 1 5 10 15

Gly Ala Gly Gly Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val
 20 25 30

Arg Ala Leu Val Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe
 35 40 45

Ser Val Ser Val Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr
 50 55 60

Tyr Ser Leu Gly Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser
 65 70 75 80

Thr Gly Val Ala Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe
 85 90 95

Cys Gly Cys His Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg
 100 105 110

Pro Leu Pro Ala Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg
 115 120 125

Tyr Arg Tyr Tyr Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp
 130 135 140

Trp Asp Trp Ala Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn
 145 150 155 160

Gly Ile Asn Leu Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln
 165 170 175

Arg Val Tyr Leu Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe
 180 185 190

Phe Thr Gly Pro Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His
 195 200 205

Thr Trp Asp Gly Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr
 210 215 220

Leu Gln His Arg Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro
 225 230 235 240

- 52 -

Val	Leu	Pro	Ala	Phe	Ala	Gly	His	Val	Pro	Glu	Ala	Val	Thr	Arg	Val	245	250	255
Phe	Pro	Gln	Val	Asn	Val	Thr	Lys	Met	Gly	Ser	Trp	Gly	His	Phe	Asn	260	265	270
Cys	Ser	Tyr	Ser	Cys	Ser	Phe	Leu	Leu	Ala	Pro	Glu	Asp	Pro	Ile	Phe	275	280	285
Pro	Ile	Ile	Gly	Ser	Leu	Phe	Leu	Arg	Glu	Leu	Ile	Lys	Glu	Phe	Gly	290	295	300
Thr	Asp	His	Ile	Tyr	Gly	Ala	Asp	Thr	Phe	Asn	Glu	Met	Gln	Pro	Pro	305	310	315 320
Ser	Ser	Glu	Pro	Ser	Tyr	Leu	Ala	Ala	Ala	Thr	Thr	Ala	Val	Tyr	Glu	325	330	335
Ala	Met	Thr	Ala	Val	Asp	Thr	Glu	Ala	Val	Trp	Leu	Leu	Gln	Gly	Trp	340	345	350
Leu	Phe	Gln	His	Gln	Pro	Gln	Phe	Trp	Gly	Pro	Ala	Gln	Ile	Arg	Ala	355	360	365
Val	Leu	Gly	Ala	Val	Pro	Arg	Gly	Arg	Leu	Leu	Val	Leu	Asp	Leu	Phe	370	375	380
Ala	Glu	Ser	Gln	Pro	Val	Tyr	Thr	Arg	Thr	Ala	Ser	Phe	Gln	Gly	Gln	385	390	395 400
Pro	Phe	Ile	Trp	Cys	Met	Leu	His	Asn	Phe	Gly	Gly	Asn	His	Gly	Leu	405	410	415
Phe	Gly	Ala	Leu	Glu	Ala	Val	Asn	Gly	Gly	Pro	Glu	Ala	Ala	Arg	Leu	420	425	430
Phe	Pro	Asn	Ser	Thr	Met	Val	Gly	Thr	Gly	Met	Ala	Pro	Glu	Gly	Ile	435	440	445
Ser	Gln	Asn	Glu	Val	Val	Tyr	Ser	Leu	Met	Ala	Glu	Leu	Gly	Trp	Arg	450	455	460
Lys	Asp	Pro	Val	Pro	Asp	Leu	Ala	Ala	Trp	Val	Thr	Ser	Phe	Ala	Ala	465	470	475 480
Arg	Arg	Tyr	Gly	Val	Ser	His	Pro	Asp	Ala	Gly	Ala	Ala	Trp	Arg	Leu	485	490	495
Leu	Leu	Arg	Ser	Val	Tyr	Asn	Cys	Ser	Gly	Glu	Ala	Cys	Arg	Gly	His	500	505	510
Asn	Arg	Ser	Pro	Leu	Val	Arg	Arg	Pro	Ser	Leu	Gln	Met	Asn	Thr	Ser	515	520	525

- 53 -

Ile Trp Tyr Asn Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu
 530 535 540

Thr Ser Ala Pro Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu
 545 550 555 560

Leu Asp Leu Thr Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr
 565 570 575

Glu Glu Ala Arg Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu
 580 585 590

Arg Ala Gly Gly Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu
 595 600 605

Val Leu Ala Ser Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln
 610 615 620

Ala Arg Ala Ala Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln
 625 630 635 640

Asn Ser Arg Tyr Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu
 645 650 655

Asp Tyr Ala Asn Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr
 660 665 670

Pro Arg Trp Arg Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln
 675 680 685

Gly Ile Pro Phe Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu
 690 695 700

Glu Gln Ala Phe Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg
 705 710 715 720

Gly Asp Thr Val Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro
 725 730 735

Gly Trp Val Ala Gly Ser Trp
 740

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 54 -

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(viii) POSITION IN GENOME:

(A) CHROMOSOME/SEGMENT: Chromosome 17

(ix) FEATURE:

(A) NAME/KEY: exon 1

(B) LOCATION: 990..1372

(ix) FEATURE:

(A) NAME/KEY: exon 2

(B) LOCATION: 2115..2262

(ix) FEATURE:

(A) NAME/KEY: exon 3

(B) LOCATION: 3056..3202

(ix) FEATURE:

(A) NAME/KEY: exon 4

(B) LOCATION: 3387..3472

(ix) FEATURE:

(A) NAME/KEY: exon 5

(B) LOCATION: 5667..5923

(ix) FEATURE:

(A) NAME/KEY: exon 6

(B) LOCATION: 7745..8955

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATAATGAGCA GTGAGGACGA TCAGAGGTCA CCTTCCTGTC TTGGTTTTGG CAGGTTTTGA	60
CCAGTTTCTT TGCTGCATTC TGTTTTATCA GCGGGGTCTT GTGACCTTTT ATCTTGTGCT	120
GACCTCCTGT CTCATCCTGT GACGAAGGCC TAACCTCCTG GGAATTCAGC CCAGCAGGTC	180
TCTGCCTCAT TTTACCCAGC CCCTGTTCAA GATGGAGTCG CTCTGGTTGG AAACCTCTGA	240
CAAAATGACA GCTCCTGTTA TGTTGCTGCT GCTGCCGCCA ATGGACAGCC TTTAACGTGC	300
CCGCCAGCCC TGCTCCACCG CCGGCCTGGG CTCACATGGC CCCATCCCTC CTCGAACCTC	360
CTAGCCTGTT AGTTACTCAA ATCTGCAAGC TCTCTGCCTT CTCAGGGCCT TCAATAAATG	420
CATTCTTTCT GTCTGGAAGG CTCTTCCTTT CCCTCTTCTA GCCAATTCCT ATTCATCCCT	480
GAGTTTCAGA TTAAGAGTCA CTTCTTTTGG AAACCTTACT TCGCTACTTC GCTACTTACT	540
GCACTACTTC GCAGCATCAC AACTATGATG GAAATCCTTA CTTACGTTAA ATATCTGGTT	600

- 55 -

TCTAGGTCAC CTCCCTGACG GGGACGGTAG GGACCGTCTT CTCGTTTCATC AGTAGGGAAG	660
TAGCTATGGC AGTGCCTGAT ACAAATAAA CTCCAAATGT GTATTTATTA GATGGTTGGA	720
TGGAAGTTAT TTGCGTGTGA AAGCGCGTTT TACCCGAAGG CGCTCTGTGA GGGCCAGCGG	780
GTCCCCCTTCG GCCCTGGAGC CGGGGTCACA CGCTCCCCAC CGCGTGCGGT CACGAGACGC	840
CCCCAAGGGA GTATCCTGGT ACCCGGAAGC CGCGACTCCT GGCCCTGAGC CCGGGCTTAG	900
CCTTCGGGTC CACGTGGCCG GAGCCGGCAG CTGATTGGAC GCGGGCCGCC CCACCCCTG	960
GCCGTCGCGG GACCCGCAGG ACTGAGACCA TGGAGGCGGT GCGGTTGCC GCGCGGTGG	1020
GGGTCTTTCT CCTGGCCGGG GCCGGGGGCG CGGCAGGCGA CGAGGCCCGG GAGGCGGCGG	1080
CCGTGCGGGC GCTCGTGGCC CGGCTGCTGG GGCCAGGCC CGCGGCCGAC TTCTCCGTGT	1140
CGGTGGAGCG CGCTCTGGCT GCCAAGCCGG GCTTGGACAC CTACAGCCTG GGCGGCGGCG	1200
GCGCGGCGCG CGTGCGGGTG CGCGGCTCCA CGGGCGTGGC GGCCGCCGCG GGGCTGCACC	1260
GCTACCTGCG CGACTTCTGT GGCTGCCACG TGGCTGGTC CGGCTCTCAG CTGCGCCTGC	1320
CGCGGCCACT GCCAGCCGTG CCGGGGGAGC TGACCGAGGC CACGCCAAC AGGTACCGCC	1380
CCGAAGCTTC CCCGCGTCCG CCCGAGGCGC TTACCCCTC CCGGAGCCGC TGCCACCCAA	1440
ATCGGGAGGC TGAGCGGGGA GCGCTGGCCG GAAGGCCCGA CTGCGCCGCC TCCAGCAGCT	1500
GTGTGGCCTT GAGCCAGCCA CTCTGCCTTT CAGAGCCTCG GCTGGCCAC CTGAAAAACG	1560
GAAAGAAGAC GCCTACCGTG CAGTGTATT GTGAGGATT GCACGATGAT GGCATAGAA	1620
TTGTGGTGC ACAATTGGTG ATGAGTGAAT TTTCTTGCT TCCTCCCCA CCTTCTCTTT	1680
GAACCTGCGG ACTGAGGAAG GACGCCTCCA TCCCCACCC TACAGGCCTG TGTCCAGCG	1740
CCTGCCACAC TATGGAGTGA TGTGTTACA CAGCTGTCCT CCCCTGCCCA TCTGTTAGAC	1800
TGTGGGGGCA GGGATTCCCC GTTCCAGGAA AACACCGTGC AGAGGAGGGG CTCTGGCAGT	1860
GTGGCATGAA AGTGGAATAT GCCACCCAAA TACCCGCCAG GCTAGAGGGC CCTGGGAGAG	1920
TGCAGGGGAC GAGTGCCTCA GAAGCCCAGC CCCGGTACCT GGTCTCAGCT CCACCTGGGG	1980
TGGGTCCCAG TGTGCAGCAG AAGGGCCGAG TTTGGAGCCC CTCCCTCTC CTCTAGGTGG	2040
GGGATGGGGG ATTTGTTCCA GGGCCGTGGA CCCTCCAGGG TGGGATGCGC CCCTGCTCAT	2100
GAACTGCCC GCAGGTACCG CTATTACCAG AATGTGTGCA CGCAAAGCTA CTCCTTCGTG	2160
TGGTGGGACT GGGCCCGCTG GGAGCGAGAG ATAGACTGGA TGGCGCTGAA TGGCATCAAC	2220

- 56 -

CTGGCACTGG CCTGGAGCGG CCAGGAGGCC ATCTGGCAGC GGGTGCCTGC CCACTGTCCC	2280
TTCCCCACCC TCCTCTATGG CGGGAGCCAC CGTAGGTGTT TTCACCCGCC CCCCAGCATG	2340
GGCGCAGTGT CTCTCTCTAG AAGTGCTTTC AGCGTGCA CA GTGGCTTGGG CCTCCTAAAA	2400
ACTGAGGCTT CCGGCCGGGC GCGGTGGCTC ACGCCTGTCA TCCCAGCACT TCGGGAGGCC	2460
TAGGCGGGCG GATCAGGAGT TCAGGAGATC GAGACCATCC TGGCCAACAT TGTGAAACCC	2520
CGTCTCTACT AAAATACAAA GAAATAGCAA CCTGGGCAAC AGAGCGAGAC TCTGTCTAAA	2580
AAAAAAAAAA AAAAAAATG AGGCTTCCAG TTTGAGGAGT GGGGCTCCTT CCCCCATCTC	2640
CCCTATGCAG CCAATCACCT GGTCCCTTGG ATCCAATCA TGGGCAGCTC TAGATCTGCC	2700
TCCCTGGAAG CTTCTGTGCT GCAATGGCTG CTCCAGGCTC TGCTTAAGCT CTTACACAG	2760
TTGCCCTGCC CTTCCATCTG GCACTCTTGC TCCATGAAGC CTTCTAAGGC CTTCTGTG	2820
GGGGAAAGCC CCTTTGTGCC CCATCTCCTC ACCCATGCGA CAAAGGCAAC ACAGTGAAC	2880
CACCTACTCA CAGGTCTCTT TCCTCTGGGC TGTGGGCTCC TTGATGGCAG CGTTCGGATT	2940
TTGTCTCAGT AGCCCTAGCA CCCAGCACAA AGAAGCAATG AGTGAATGGT TGTGAATGA	3000
ATGAATGAAT GAATGAAGAT GAATATATTT CTATGTGTGG GCCCTTCTTC CTCAGGTGTA	3060
CCTGGCCTTG GGCCTGACCC AGGCAGAGAT CAATGAGTTC TTTACTGGTC CTGCCTTCCT	3120
GGCCTGGGGG CGAATGGGCA ACCTGCACAC CTGGGATGGC CCCCTGCCCC CCTCCTGGCA	3180
CATCAAGCAG CTTTACCTGC AGGTAAAAGG ATGGAAAAGG GAAGGGGCAG AATCGGTGAT	3240
AGATGGTCAT GGGCCCAGGA AGGGTGGTAT TAGGCCGGCC CCAGGGCTCT TAACTGAGGC	3300
GGGGGGCTGC GTGTATCCTG GGAGATGAGG GCCTTCTCAT AGGACAGCAG TGGCCATGCT	3360
CACCACCCTT CTTTCTGTTC CTCCAGCACC GGGTCTGGA CCAGATGCGC TCCTTCGGCA	3420
TGACCCCAGT GCTGCCTGCA TTCGCGGGC ATGTTCCCGA GGCTGTCACC AGGTGAGGTT	3480
CCGCTCACCC CCTCCACTTA GCTCAGAGAG GGAATTTTAT TCCCTTCTAG AACATGACTT	3540
AAAAACTTAA GCTCTGGGCC GGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA	3600
GGCCGAGTTG GGCGGATCAC CTGAGGTCAG GAGTTGAGA CCAGCCTGGC CAACATGGTG	3660
AAACCCTGTC TCTACTAAAA ATATAAAAAT TAGCTGGGCA TGGTGGCACG CGCCTGTAAT	3720
CCCATCTACT TAGGAGGCTG AGACAGGAGA ATTGCTTAAA CCTGGGAGGC AGACGTTGCA	3780
GTGAGTCAAG ATCACGCCAT TGCCTCCAG CCTGGGTGAC GAGCGAACT CTGTCTCAAA	3840

- 57 -

CAAACAAACA AGCTCTGGAC GTAGGCCTGG GTTTGATTTC TGA CTCTGCT ACTAATTAGC	3900
TGTGTGACTT CGGGCAGATG ACATGACTGC TCTGTGCCTC AGTTTCCTTA CTGTAAAAAT	3960
GGGATCTCTA CCCACTTCGC TGTAGGGTTT GTAATTATCT CTCGATCTAT CTGTGACTTT	4020
GCACAGAGTG CTAGCAAATG GCAGCCCTTG GGAGTGGCAG CAGGGGTGCT CCAGTGTCCT	4080
TTGTCCCTCC TGTTCTCTG TGCTTCCCAG CCATCCTCTC ACATGTGGTT GGGAAAAGTC	4140
TTCAAGGCTC ACCTGAGACC TCCCCTCCTT CAGGAAGCCT TGCTAGTGCC CCGCATGACC	4200
TCCTTTGCAC CTGCTAATGT CTGGCTCCCA TACTCTCGTA GGACTTAATG CATGCCAGTG	4260
GCCTCCCTGC CCGCCTCTT GCCCCCATCA CCAGGTGGCA GGAAACTCAC TCATTCATTC	4320
AATAAACTTG GTCCAGCTGT CTGAGGCTGC CAGAACTGGC TGTGCTGGGT CCTGGGAGGC	4380
GGCAAGAAAG GTGCCCAAGG GCTTACCCCT GATAGGAGAG ATATGTTGGC TGAAGGATAC	4440
AATGTGGGGA CAAGGACAGG AATATATGTG GGTTCGCTC TCCTCTGCCG GGAGAGAGGG	4500
GCAGGAAGGG CTCAGGGCAG AGCCCAGCCT TGAAAAATGA GTGTTGCTTG GACGGACGCT	4560
TGGCTAATGC TTGTAATCCT AGCGTTTTGG GAGGCTGAGG CGTATGGATC ACCTGCGGTC	4620
AGGAGTTAAA GACCAGCCTG GCCAACATGG CGAAACCCCA TCTCTACTAA AAGTACAAAA	4680
ATTAGCCAGG CGTGGTGGCG GGCTCCTGTA ATCCCAGCTA CTCGGTAGGC TGAGGCATGA	4740
GAATCTCTTG AAGCCAGGGG CCAGAGACTG CAGTGAGCCG AGATCACACC ACTTCACTCC	4800
AGCCTGGGTG ACAGAGTGAG ACTCCGTCTC AAAAAAAAAA AAAAAAAAAAG GAAAGAAAAT	4860
TAAACACCTC ATGTTCTCAC TCATAGTGGG AGTTGAACAA TGAGAACAAC ATGGACACAG	4920
GAAGGGGAAC ATCACACACC GGGGCCTTTC GCGGTGTGGG GGTCAAGGGG AGGAGTAGCA	4980
TTGGGACAGA TACTTAATGC ATGCGGGGCT GAAAACCTAG ATGATGGGTT GATGGGTGCA	5040
GCAAACCACC ATGGCACATG TATACCTATG CAACAAACCT GCATGTTCTG CACAGAACTG	5100
AACTGAAAGT ATAATTAAAA AAAAAAAAAA AAGCTGGGTG CGGTGGCCCA CACCTGTAAT	5160
CCCAGCACTT TGGGAGGCCG AGACGGGCGG ATCACAAGGT CAGCAGATCG AGACCATCCT	5220
GGCTAACACA GTGAAACTCA GTCTCTACTA AAAATACAAA AAATTAGCCG GGTGTGGTGG	5280
CGGGCACCTG TAGTCCCAGC TACTAGGGAG GCTGAGGCAG GAGAATGGCA TGAACCTGGG	5340
AGGCAGAGCT TGCAGTGAGC TGAGAATGCG CCACTGCACT CCAGCCTGGG GGACAGAGTG	5400
AGACTCTGCC TCAAAAAAAAAA AAAAAAAAAAG AAAGAAAAAG GAGCGTTGCT TGTTCAGGC	5460

- 58 -

CACAGGAAGG GGAGAGATAG TGAAAGTTTT TCAGAGAAGG TGGCCAGGGA AGGAGAAGAA	5520
AGGACTGTAG GCAGAGAGCA TAGCCTGTAC AAAGCCATAG AGGCAAGAGA AACCAGGAGC	5580
TGTAGAGAAG TTGGCAAGGC TGTGTAACAC TATGGTGAAC ACTATGGCGG CTTCATGAA	5640
ATATCTGAGC TTTTGCTCCC CACTAGGGTG TTCCCTCAGG TCAATGTCAC GAAGATGGGC	5700
AGTTGGGGCC ACTTTAACTG TTCCTACTCC TGCTCCTTCC TTCTGGCTCC GGAAGACCCC	5760
ATATTCCCCA TCATCGGGAG CCTCTTCCTG CGAGAGCTGA TCAAAGAGTT TGGCACAGAC	5820
CACATCTATG GGGCCGACAC TTTCAATGAG ATGCAGCCAC CTTCTCAGA GCCCTCCTAC	5880
CTTGCCGCAG CCACCACTGC CGTCTATGAG GCCATGACTG CAGGTACAGT GCCTGGGTGG	5940
GGTGGGAGAG CCCCCAGAC CCTCAAAAAG AAGGGAGTAG CAGATGTCAG TAGGGGTAGG	6000
CAGAGGGACT GGAATAATGC CTCGCCATAA CACACAGTAC TTTATAGTTT ACCAAGCACG	6060
TGTACACATG CGTTGTCTCA GTGAATCCCA CTGTGGTTGA GAGGTGAGCT CTGGAAGCCA	6120
ACAACCTGGG TCACACCTCG CGCTCCTATT TCCTGGCCGT GTGACTTATG ACTCATGACC	6180
TCCTTCCCAG TGTCTCGTTT GCTTTTCTCG TAAACTGGGA CTACCTCATA GGTAAGATAA	6240
CGCCTGGCCC AGAGCAAAGG CCACTAAGAG CTAGCTATGA ACAAGGATTT TGTTTCATCT	6300
CTGCGTGGTT GCTGAAGTAG GCACTGCAGG CAGGAGGTGA GTGGATGTGC CTAAAGGCAC	6360
TAAGTGCGCA TCCTGCTACA AAAGTGTGAA GCCAGGGCTC CTTCTGCCA CTTAAAGGAG	6420
GAGTGGAGCA GAGGGCGCCC AAGTCAGGAA TGACTTAGTG GAGAGGCGTC TGTGTTGGCC	6480
AGGAAGGGAA CAGATCAGCT CAGCCTTTCT TGAGCAGTAC TGCTCCAAGT GTGACCCAAA	6540
ACCAGCAGCA GCAGCAGCAG CAGCCCGAGC TGTGAGATGG CAAATTCTCA GGCCCTACCC	6600
AAGACCTGAA GGAGAAGCTA CATTTTTTTT TTTTTTGAGA CAGATTTCAC TCTGTTGCTG	6660
AGGCTGGAGC ACAGTGGCAC AATCTCATCT CACTGCAACC TTCGTCTCCT AGGTTCAAGC	6720
GATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTATAG GCACCCGCCA CCACGCCCGG	6780
CAATTTTTGT TTGTTTTGAG ATAGAGTCTC GCTCTGTCAC CCAGGCTGGA GTGCAGTGGC	6840
ACGATCTCAG TTTCACTGCAA CCTCTGCTTC CTGAGTTCAA GCGATTCTCC TGCTCAGCC	6900
TCCTGAGTAG CTGGGATTAC AGGCGCCCCC CAACCACACT CGGCTAATTT TTGTATTTTT	6960
AGTAGAGACG GGGTTTCGCT ATGTAGGTCA AGCTGGTTTC AAAGTCTGA CCTCAAATGA	7020
TTGCCCCACT TCAGCCTCCC AAAGTGCTGG GATTACAGGT GTGAGCCACC TTGCTGGCC	7080

- 59 -

AATTTTGTGA TTTTGTAG AAACAGGTTT CACCATGGTG GCCAGACTGG TCTCAAATC	7140
CTGACCTCAG GTGAACTGCC CACCTCAGCC TCCCAAAGTA CTGGTATTAC AGGCGTGATC	7200
CACTGCGACT GGCCTTGATT TTGTTTTTGA GACAGAATCT TACTCTGTCT CCCAGACTGG	7260
AGTGCACTGG CACAATCTCA GCTCACTGCA ACTTCTGCCT CATGGGTTCAGT GATTCTT	7320
GTGCTCTTAC CTCCCGAGTA GCCGGGATTA CAGGCACCTG CCATTACGCT AGGCTAATTT	7380
TTGTATTTTT AGTATAGACA GGGTTTCCCC ACATTGGCCA GGCTGGTCTG GAACTCCTGG	7440
GCTCAAGTGA TCCACCTGCT TCAGCCCTC AGAGTACTGG GATTATAGGT GTGGGCCACC	7500
ACGCCCCATT AGAAACCTCC ATGTTTTAAG GAGCCCTCTG GGTAACCTC ATGTTACCC	7560
AAGCTGCTGA ACCCTGTCCT GGAGTTTCA GAGGGACGCG TATGTGCCAC AGAGCGTCCC	7620
GCTGGTGGGG GTCATGGGAA GCCATGACCT GGGATAGACA GTCGTCTGTA GAGTGGGGTG	7680
AACATTCCCT GGGCCCTCTG TTTCATCACT CCTCTTCTCT GTTCCCCCTA CCTCCTGTCC	7740
ACAGTGATA CTGAGGCTGT GTGGCTGCTC CAAGGCTGGC TCTTCCAGCA CCAGCCGAG	7800
TTCTGGGGGC CCGCCAGAT CAGGGCTGTG CTGGGAGCTG TGCCCCGTGG CCGCCTCTG	7860
GTTCTGGACC TGTTTGCTGA GAGCCAGCCT GTGTATACCC GCACTGCCTC CTCCAGGGC	7920
CAGCCCTTCA TCTGGTGCAT GCTGCACAAC TTTGGGGGAA ACCATGGTCT TTTGGAGCC	7980
CTAGAGGCTG TGAACGGAGG CCCAGAAGCT GCCCGCCTCT TCCCCAATC CACCATGGTA	8040
GGCACGGGCA TGGCCCCCGA GGGCATCAGC CAGAACGAAG TGGTCTATTC CCTCATGGCT	8100
GAGCTGGGCT GCGGAAAGGA CCCAGTGCCA GATTTGGCAG CCTGGGTGAC CAGCTTTGCC	8160
GCCCCGCGGT ATGGGGTCTC CCACCCGGAC GCAGGGGCAG CGTGGAGGCT ACTGCTCCGG	8220
AGTGTGTACA ACTGCTCCGG GGAGGCCTGC AGGGGCCACA ATCGTAGCCC GCTGGTCAGG	8280
CGGCCGTCCC TACAGATGAA TACCAGCATC TGGTACAACC GATCTGATGT GTTTGAGGCC	8340
TGGCGGCTGC TGCTCACATC TGCTCCCTCC CTGGCCACCA GCCCCGCTT CCGCTACGAC	8400
CTGCTGGACC TCACTCGGCA GGCAGTGCA GAGCTGGTCA GCTTGACTA TGAGGAGGCA	8460
AGAAGCGCCT ACCTGAGCAA GGAGCTGGCC TCCCTGTTGA GGGCTGGAGG CGTCTGGCC	8520
TATGAGCTGC TGCCGGCACT GGACGAGGTG CTGGCTAGTG ACAGCCGCTT CTGCTGGGC	8580
AGCTGGCTAG AGCAGGCCCC AGCAGCGGCA GTCAGTGAGG CCGAGGCCGA TTTCTACGAG	8640
CAGAACAGCC GCTACCAGCT GACCTTGTGG GGGCCAGAAG GCAACATCCT GGACTATGCC	8700

- 60 -

AACAAGCAGC TGGCGGGGTT GGTGGCCAAC TACTACACCC CTCGCTGGCG GCTTTTCCTG	8760
GAGGCGCTGG TTGACAGTGT GGCCCAAGGC ATCCCTTTCC AACAGCACCA GTTTGACAAA	8820
AATGTCTTCC AACTGGAGCA GGCCTTCGTT CTCAGCAAGC AGAGGTACCC CAGCCAGCCG	8880
CGAGGAGACA CTGTGGACCT GGCCAAGAAG ATCTTCCTCA AATATTACCC CGGCTGGGTG	8940
GCCGGCTCTT GGTGATAGAT TCGCCACCAC TGGGCCTTGT TTTCCGCTAA TTCCAGGGCA	9000
GATTCCAGGG CCCAGAGCTG GACAGACATC ACAGGATAAC CCAGGCCTGG GAGGAGGCC	9060
CACGGCCTGC TGGTGGGGTC TGACCTGGGG GGATTGGAGG GAAATGACCT GCCCTCCACC	9120
ACCACCCAAA GTGTGGGATT AAAGTACTGT TTTCTTTCCA CTAAACTGA TGAGTCCCCT	9180
GGGTCTGTCA AAATGAGAAG GTCACTGCTG CCACGCTTGG GAGGACTCAG GGCTATAGCA	9240
TGGCCCTGGG GTGGGACCTG TTCTCCCATC CCTTGCCTCA CGTCCCTGTT TTTGTTTGT	9300
TGTTTGTGTTG TGACGGAGCC TTGGTCTGTT GCCCAGGCTT GAGTACAATG GCACAGTCTC	9360
GGCTCACTGC AACCTCCGCC TCCTGGGTTC AAGCAATTCT TGTGCCTCAG CCTCCCCGGT	9420
AGCTGGGACT ATAGGCATGC ACCACCACAC CAGGCTAATT TTTTTTTTC CAAGATGGAG	9480
TCTTGCTCTG TCGCCAGGT TGGAGTTAG TGGCACCATA TTGGTTTACT GCAACCTCTG	9540
CCTCCCCGGT TCAAGCAATT CTCCTGCCTC AGTCTACCAG GGAGTTAGGA CTACGGGCCT	9600
GTGCCATCAC GCCTGGCTAA TTTTGTATT TTTCATAGAG ATAAGGTTTC ACCATGTGG	9660
CCAGGCTGGT CTTAACTCC TGAAGTCAAG TGATCCACCT GCCTCGGCCT TCCAAAGTGC	9720
TGGGATTACA GGAGTGAGCC ACCGTGCCC GGCATGTCTC TCTTTTAACT ACTAATGTTA	9780
CCCTGACCTT TGAACGTAGA ATGCCCTTCT GTTGCAGGAA AACCTCTTTT CAAACCATGT	9840
TTGTCTTTG CTGGCATGCC ACAGCAACAG TCACCAACAC AGAAGACTTC TGTGACCAA	9900
TATTTGGAGG ATTTTCCCCA CACACACCAA GCAGCAGACA TCAGCTGGGT GTCCTCCAAT	9960
TCAGTTCCAA TGAATCAAC CAGAGACAGC ATCAGATCCC ACAGGGTTAG GGTGCAGATC	10020
CATGAGACCA CCCCCTCCTT CCCAACGGTT ACAAGTCTG ATCCCTGGAA CTTCTGACTA	10080
ACTGGCTTCA AGTTGGAGTT CCCATGACCC CTTTCCCCTC TTTGGAGTCA ACTCATTTGC	10140
GACAGTGACC CACGAAACAC AGGGAAACCC TTATTATGTT TATTGCTTTA TTACAGAGGA	10200
AAAAAATTTT TTTCTTTCTT TTTGAGACA GGGTCTCACT CTGTCATCCA GAATGACTGC	10260
AGTGGCAGGA TCTGGCTCCG TCACCCAGGC TGGAGTGCAG TGGCATGATC TCGGCTCACT	10320

- 61 -

ACAGCCTCCA TCCCCCCCAA ACCCCACGCC TCAGCGCCCC ACCCCGCAAG TGGCTGGGAC 10380

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp	Glu	Ala	Arg	Glu	Ala	Ala	Ala	Val	Arg	Ala	Leu	Val	Ala	Arg	Leu
1				5					10					15	
Leu Gly Pro Gly															
20															

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

- 62 -

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: Modified-site, glycosylated or phosphorylated,
wherein Xaa may be any amino acid residue, preferably Arg.
(B) LOCATION: 16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly Gly Gly Ala Ala Xaa Val
1 5 10 15

Arg

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: Modified-site, glycosylated or phosphorylated,
wherein Xaa may be any amino acid residue, preferably Ala
(B) LOCATION: 12

(ix) FEATURE:

- (A) NAME/KEY: Modified-site, glycosylated or phosphorylated,
wherein Xaa may be any amino acid residue, preferably Ser
(B) LOCATION: 14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Trp Arg Leu Leu Leu Thr Ser Ala Pro Ser Leu Xaa Thr Xaa Pro
1 5 10 15

CLAIMS:

1. An isolated nucleic acid molecule comprising a sequence of nucleotides which encodes or is complementary to a sequence which encodes a mammalian α -N-acetylglucosaminidase or fragment or derivative thereof.
2. The isolated nucleic acid molecule according to claim 1 wherein the nucleotides are deoxyribonucleotides.
3. The isolated nucleic acid molecule according to claim 2 wherein said molecule is cDNA.
4. The isolated nucleic acid molecule according to claim 2 wherein said molecule is a genomic DNA molecule.
5. The isolated nucleic acid according to any one of claims 1 to 4 wherein the mammal is a human.
6. The isolated nucleic acid according to claim 5 wherein the α -N-acetylglucosaminidase is of liver, kidney or placenta origin.
7. The isolated nucleic acid molecule according to any one of claims 1 to 6 having a nucleotide sequence substantially as set forth in SEQ ID NO:1 or complementary thereto or having at least 40% similarity to all or part thereof.
8. The isolated nucleic acid molecule according to any one of claims 1 to 6 having a nucleotide sequence substantially as set forth in SEQ ID No:3 or complementary thereto or having at least 40% similarity to all or part thereof.
9. The isolated nucleic acid molecule according to claim 7 or 8 wherein the percentage similarity is at least 60%.

- 64 -

10. The isolated nucleic acid molecule according to claim 9 wherein the percentage homology is at least 80%.
11. The isolated nucleic acid molecule according to any one of claims 1 to 10 wherein the α -N-acetylglucosaminidase or fragment or derivative thereof encoded by said molecule comprises an amino acid sequence substantially identical to SEQ ID No:2 or is at least 40% similar to all or a part thereof.
12. The isolated nucleic acid molecule according to claim 11 wherein the percentage similarity to SEQ ID No:2 is at least 60%.
13. The isolated nucleic acid molecule according to claim 12 wherein the percentage similarity to SEQ ID No:2 is at least 80%.
14. The isolated nucleic acid molecule according to any one of claims 1 or 13 wherein said molecule is carried by a vector capable of replication in a eukaryotic cell and/or a prokaryotic cell.
15. The isolated nucleic acid molecule according to claim 14 wherein the vector is an expression vector.
16. The isolated nucleic acid molecule according to claim 15 wherein the expression vector is capable of being expressed in cells derived from a eukaryote.
17. The isolated nucleic acid molecule according to claim 16 wherein the expression vector is further capable of being expressed in cells derived from a mammal.
18. The isolated nucleic acid molecule according to claim 17 wherein the expression vector is further capable of being expressed in CHO cells.

- 65 -

19. A recombinant mammalian α -N-acetylglucosaminidase or fragment or derivative thereof.
20. The recombinant mammalian α -N-acetylglucosaminidase according to claim 19 in substantially pure form.
21. The recombinant mammalian α -N-acetylglucosaminidase according to claims 19 or 20 when expressed in mammalian, yeast or insect cells.
22. The recombinant mammalian α -N-acetylglucosaminidase according to claim 21 when expressed in mammalian cells.
23. The recombinant mammalian α -N-acetylglucosaminidase according to claims 21 or 22, wherein the cells are capable of glycosylating said recombinant mammalian α -N-acetylglucosaminidase.
24. The recombinant mammalian α -N-acetylglucosaminidase according to claim 23 wherein the cells are capable of N-glycosylating said recombinant mammalian α -N-acetylglucosaminidase.
25. The recombinant mammalian α -N-acetylglucosaminidase according to claim 24 wherein the cells are CHO cells.
26. The recombinant mammalian α -N-acetylglucosaminidase according to any one of claims 19 to 25 wherein said recombinant α -N-acetylglucosaminidase is in a glycosylated form.
27. The recombinant mammalian α -N-acetylglucosaminidase according to claim 26 wherein the molecular weight of the glycosylated form as determined using SDS/PAGE is at least approximately 79 kDa.

- 66 -

28. The recombinant α -N-acetylglucosaminidase wherein the molecular weight of the glycosylated form as determined using SDS/PAGE is at least approximately 79 kDa to 89 kDa.
29. The recombinant mammalian α -N-acetylglucosaminidase according to any one of claims 19 to 28 comprising a sequence of amino acids substantially the same as a human α -N-acetylglucosaminidase.
30. The recombinant mammalian α -N-acetylglucosaminidase according to any one of claims 19 to 29 when fused to another proteinaceous molecule.
31. The recombinant mammalian α -N-acetylglucosaminidase according to claim 30 wherein the other proteinaceous molecule is an enzyme, reporter molecule, purification site and/or a signal sequence.
32. The recombinant mammalian α -N-acetylglucosaminidase according to any one of claims 19 to 31 comprising an amino acid sequence substantially as set forth in SEQ ID NO:2 or having at least 40% similarity to all or part thereof.
33. The recombinant mammalian α -N-acetylglucosaminidase according to claim 32 wherein the percentage similarity to SEQ ID NO:2 is at least 60%.
34. The recombinant mammalian α -N-acetylglucosaminidase according to claim 33 wherein the percentage similarity to SEQ ID NO:2 is at least 80%.
35. A recombinant α -N-acetylglucosaminidase when produced by expression of a nucleic acid molecule according to any one of claims 14 to 18.
36. The recombinant α -N-acetylglucosaminidase according to claim 35 when glycosylated.

- 67 -

37. A method of diagnosing a mutation in a gene which encodes α -N-acetylglucosaminidase in a human patient said method comprising contacting genomic DNA or RNA derived from said patient with one or more isolated DNA molecules or oligonucleotides comprising at least 10 contiguous nucleotides derived from SEQ ID No:1 or SEQ ID No:3 or a complementary strand thereof for a time and under conditions sufficient for hybridisation to occur and then detecting said hybridisation using a detection means.

38. The method according to claim 37 wherein the detection means is a reporter molecule covalently attached to the isolated DNA molecule or oligonucleotide.

39. The method according to claim 38 wherein the reporter molecule is ^{32}P , ^{35}S or biotin.

40. The method according to claim 37 wherein the detection means is a polymerase chain reaction format.

41. The method according to claim 40 wherein the polymerase chain reaction format is selected from the list comprising SSCP, AMD, AFLP, IRS-PCR, iPCR or RT-PCR, amongst others.

42. The method according to claim 41 wherein the polymerase chain reaction format is SSCP.

43. The method according to any one of claims 37 to 42 wherein the isolated DNA molecule or oligonucleotide comprises at least 20 contiguous nucleotides derived from SEQ ID No:1 or SEQ ID No:3 or a complementary strand thereof.

44. The method according to claim 43 wherein the isolated DNA molecule or oligonucleotide comprises at least 50 contiguous nucleotides derived from SEQ ID No:1 or SEQ ID No:3 or a complementary strand thereof.

- 68 -

45. The method according to claim 44 wherein the isolated DNA molecule or oligonucleotide comprises at least 100 contiguous nucleotides derived from SEQ ID No:1 or SEQ ID No:3 or a complementary strand thereof.

46. A method for treating a patient suffering from α -N-acetylglucosaminidase deficiency said method comprising administering to said patient an effective amount of recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof.

47. The method according to claim 46 wherein the mammalian α -N-acetylglucosaminidase comprises a sequence of amino acids substantially the same as the amino acid sequence of human α -N-acetylglucosaminidase.

48. The method according to claim 47 wherein the patient is suffering from mucopolysaccharidosis type IIIB.

49. The method according to any one of claims 46 to 48 wherein the recombinant α -N-acetylglucosaminidase is produced in mammalian cells.

50. The method according to claim 49 wherein the mammalian cells are capable of glycosylating the recombinant α -N-acetylglucosaminidase produced therein.

51. The method according to claim 50 wherein the recombinant α -N-acetylglucosaminidase is in a glycosylated form.

52. The method according to claim 51 wherein the glycosylated form of the recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of at least approximately 79kDa.

53. The method according to claim 52 wherein the glycosylated form of the

- 69 -

recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of approximately 79 kDa to 89 kDa.

54. The method according to anyone of claims 46 to 53 wherein the recombinant α -N-acetylglucosaminidase comprises a sequence of amino acids substantially as set forth in SEQ ID No:2 or having at least 40% similarity to all or a part thereof.

55. The method according to claim 54 wherein the percentage similarity to SEQ ID NO:2 is at least 60%.

56. The method according to claim 55 wherein the percentage similarity to SEQ ID NO:2 is at least 80%.

57. A method for treating a patient suffering from α -N-acetylglucosaminidase deficiency said method comprising administering to said patient an effective amount of recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof, wherein said recombinant mammalian α -N-acetylglucosaminidase is produced by expression of a nucleic acid molecule according to any one of claims 14 to 18.

58. The method according to any one of claims 46 to 57 wherein administration of the recombinant mammalian α -N-acetylglucosaminidase is by oral, intravenous, suppository, intraperitoneal, intramuscular, intranasal, intradermal or subcutaneous administration by infusion or implantation or by enzyme replacement therapy or by gene therapy.

59. The method according to claim 58 wherein the method of administration is by enzyme replacement therapy.

60. A pharmaceutical composition comprising a recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof and one or more pharmaceutically acceptable carriers and/or diluents.

- 70 -

61. The pharmaceutical composition according to claim 60 wherein the recombinant mammalian α -N-acetylglucosaminidase comprises a sequence of amino acids substantially the same as human α -N-acetylglucosaminidase.

62. The pharmaceutical composition according to claims 60 or 61 wherein the recombinant mammalian α -N-acetylglucosaminidase is produced in a mammalian cell.

63. The pharmaceutical composition according to claim 62 wherein the mammalian cell is a CHO cell line which is capable of glycosylating the recombinant mammalian α -N-acetylglucosaminidase.

64. The pharmaceutical composition according to any one of claims 60 to 63 wherein the α -N-acetylglucosaminidase is glycosylated.

65. The pharmaceutical composition according to claim 64 wherein the recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of at least approximately 79 kDa.

66. The pharmaceutical composition according to claim 65 wherein the recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of approximately 79kDa to 89 kDa.

67. The pharmaceutical composition according to any one of claims 60 to 66 wherein the recombinant α -N-acetylglucosaminidase comprises a sequence of amino acids substantially as set forth in SEQ ID No:2 or having at least 40% similarity to all or part thereof.

68. The pharmaceutical composition according to claim 67 wherein the percentage similarity to SEQ ID NO:2 is at least 60%.

69. The pharmaceutical composition according to claim 68 wherein the percentage

similarity to SEQ ID NO:2 is at least 80%.

70. A pharmaceutical composition comprising recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof and one or more pharmaceutically acceptable carriers and/or diluents wherein said recombinant mammalian α -N-acetylglucosaminidase is produced by expression of a nucleic acid molecule according to any one of claims 14 to 18.

71. A pharmaceutical composition comprising recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof and one or more pharmaceutically acceptable carriers and/or diluents when used in the method according to any one of claims 46 to 59.

72. Use of recombinant mammalian α -N-acetylglucosaminidase or an active fragment or derivative thereof in the manufacture of a medicament for the treatment of α -N-acetylglucosaminidase deficiency in a patient.

73. The use according to claim 72 wherein the recombinant mammalian α -N-acetylglucosaminidase comprises a sequence of amino acids substantially the same as the amino acid sequence of human α -N-acetylglucosaminidase.

74. The use according to claims 72 or 73 wherein the patient is suffering from mucopolysaccharidosis type IIIB.

75. The use according to claim 74 wherein the recombinant α -N-acetylglucosaminidase is expressed in mammalian cells.

76. The use according to claim 61 wherein the cells are CHO cells.

77. The use according to any one of claims 72 to 76 wherein the α -N-acetylglucosaminidase is glycosylated.

- 72 -

78. The use according to claim 77 wherein the recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of at least approximately 79kDa.

79. The use according to claim 78 wherein the recombinant α -N-acetylglucosaminidase has a molecular weight as determined using SDS/PAGE of approximately 79 kDa to 89 kDa.

80. The use according to anyone of claims 72 to 79 wherein the recombinant of α -N-acetylglucosaminidase comprises a sequence of amino acids substantially as set forth in SEQ ID No:2 or has at least 40% similarity to all or a part thereof.

81. The use according to claim 80 wherein the percentage similarity to SEQ ID NO:2 is at least 60%.

82. The use according to claim 80 wherein the percentage similarity to SEQ ID NO:2 is at least 80%.

83. A nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a polypeptide capable of hydrolysing the terminal α -N-acetylglucosaminidase residues present at the non-reducing terminus of fragments of heparan sulphate and heparin and wherein said nucleotide sequence is capable of hybridising under at least low stringency conditions to the nucleotide sequence set forth in SEQ ID NO:1.

84. A nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a polypeptide capable of hydrolysing the terminal α -N-acetylglucosaminidase residues present at the non-reducing terminus of fragments of heparan sulphate and heparin and wherein said nucleotide sequence is capable of hybridising under at least low stringency conditions to the nucleotide sequence set forth in SEQ ID NO: 3.

- 73 -

85. A recombinant polypeptide comprising a sequence of amino acids corresponding to the amino sequence set forth in SEQ ID NO:2 or having at least 40% similarity thereto and encoded by a nucleic acid molecule which is capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3 under at least low stringency conditions.

86. A genetic construct comprising the nucleic acid molecule according to any one of claims 1 to 18 or claims 83 to 84, operably connected in the sense orientation to a promoter sequence such that said genetic construct is capable of being expressed in a eukaryotic or prokaryotic cell to produce a recombinant mammalian α -N-acetylglucosaminidase or a fragment or derivative thereof.

87. The genetic construct according to claim 86 wherein the promoter is capable of regulating expression of the recombinant α -N-acetylglucosaminidase in a mammalian cell.

88. The genetic construct according to claim 87 wherein the promoter is the CMV promoter sequence or a promoter derived therefrom.

89. The genetic construct according to any one of claims 86 to 88 further comprising a transcription terminator sequence.

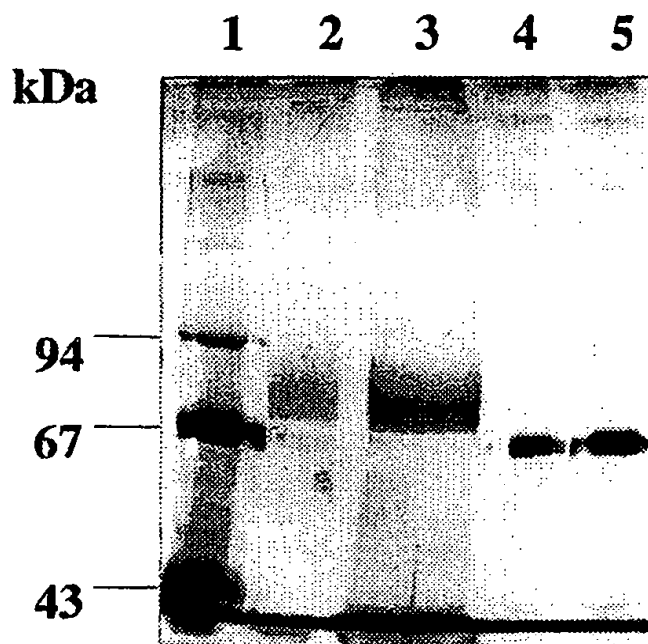
90. The genetic construct according to any one of claims 86 to 89 when used to express or over-express α -N-acetylglucosaminidase in a eukaryotic or prokaryotic cell.

91. An antibody to α -N-acetylglucosaminidase or a recombinant α -N-acetylglucosaminidase according to any one of claims 19 to 36 or an antigenic fragment thereof.

92. The antibody according to claim 91 further defined as a polyclonal antibody molecule.

- 74 -

93. The antibody according to claim 92 further defined as a monoclonal antibody molecule.

Fig. 1

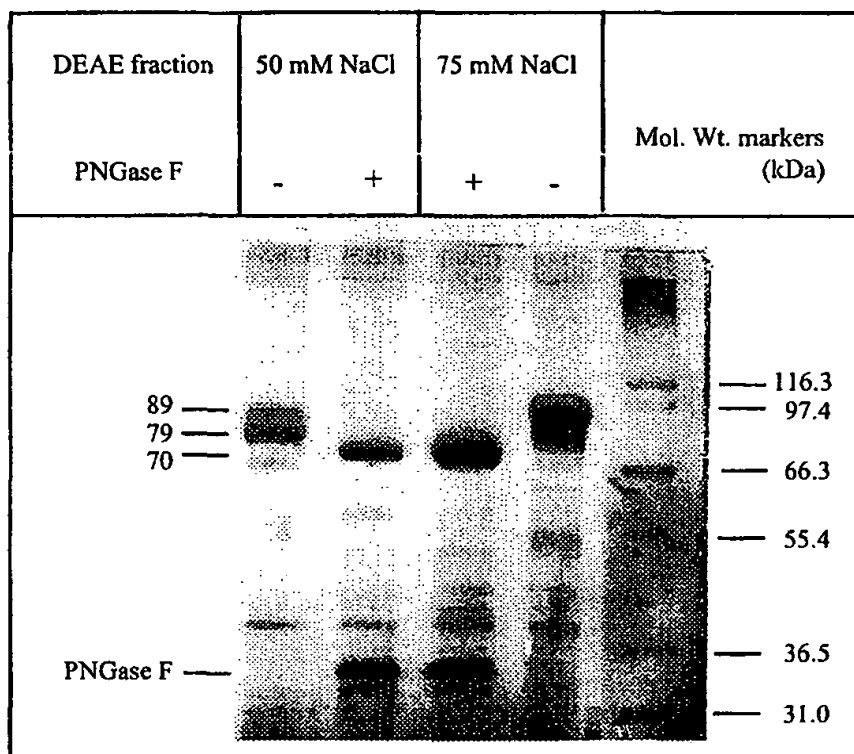


FIGURE 2

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 96/00747

A. CLASSIFICATION OF SUBJECT MATTERInt Cl⁶: C12N 15/56, 9/24; C12Q 1/68; C07K 16/40 A61K 38/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IC6: C12N 15/-: Keywords as below

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Medline: see keywords below

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
(see attached)**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	Proc. Nat. Acad. Sci. Vol. 93, pages 6101-6105 (June 1996) H. G. Zhao <u>et al</u> "The Molecular Basis of Sanfilippo Syndrome Type B" (see whole document)	1-45, 83-90
P,X	Human Molecular Genetics, Vol. 5, No. 6 pages 771-777 (1996), B. Weber <u>et al</u> : "cloning and expression of the gene involved in san Filippo B Syndrome (mucopolysacc-haridosis 111B)" (see whole document)	1-93



Further documents are listed in the continuation of Box C



See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
10 February 1997

Date of mailing of the international search report

17 FEB 1997

Name and mailing address of the ISA/AU
AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION
PO BOX 200
WODEN ACT 2606
AUSTRALIA Facsimile No.: (06) 285 3929

Authorized officer

KAREN AYERS

Telephone No.: (06) 283 2082

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 96/00747

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<u>X</u> Y	American Journal of Genetics (1994) Volume 55, A252 (Abstract 1473) H.G. Zhao <u>et al</u> : "Sanfilippo Syndrome Type B cDNA and gene encoding human α -N-acetylglucosaminidase" (see whole document)	<u>1-45, 83-90</u> 46-82
<u>X</u> Y	American Journal of Genetics (1995) Volume 57, A185 (Abstract 1059) H.G. Zhao <u>et al</u> "The Gene encoding α -N-acetylglucosaminidase and mutations underlying Sanfilippo B Syndrome" (see whole document)	<u>1-45, 83-90</u> 46-82
Y	Z. Klin. Chem. Klin. Biochem. Volume 13, pages 285-9 (1975), K. Von Figura <u>et al</u> "Serum α -N-acetylglucosaminidase: Determination, Characterization, and Corrective Activity in Sanfilippo B fibroblasts". (see whole document)	46-82
X	Eur. J. Biochem. (1976), Volume 61, Number 2, pages 581-8, K. Von Figura and H. Kresse "Sanfilippo disease type B: presence of material cross-reacting with antibodies against alpha-N-acetylglucosaminidase" (see whole document)	91-93
<u>X</u> Y	J. Biochem. (1991) Volume 110, pages 842-846, T. Sasaki <u>et al</u> "A purification and partial characterisation of α -N-acetylglucosaminidase from human liver". (see whole document)	<u>91-93</u> 1-90

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/AU 96/00747

Box B Electronic data base continued

STN: nucleotide and amino acid sequence search

WPAT/USPM/JAPIO: SS1 acetylglucosaminidase or acetylglucosaminohydrolase and
SS2: C12N 15/1C or San Filippo or mucopolysaccharos:

CAS online: SS1: as above
SS2: 3/CC or 3/SX or San Filippo or mucopolysaccharos:

medline: SS1: as above
SS2: San Filippo or mucopolysaccharos:

